

Active Calculus

Active Calculus

Matthew Boelkins
Grand Valley State University

Contributing Authors

David Austin
Grand Valley State University
Steven Schlicker
Grand Valley State University

Production Editor

Mitchel T. Keller
Morningside College

June 14, 2019

Cover Photo: James Haefner Photography

Edition: 2018

Website: <http://activecalculus.org>

© 2012–2018 Matthew Boelkins

Permission is granted to copy and (re)distribute this material in any format and/or adapt it (even commercially) under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License. The work may be used for free in any way by any party so long as attribution is given to the author(s) and if the material is modified, the resulting contributions are distributed under the same license as this original. All trademarks™ are the registered® marks of their respective owners. The graphic



that may appear in other locations in the text shows that the work is licensed with the Creative Commons and that the work may be used for free by any party so long as attribution is given to the author(s) and if the material is modified, the resulting contributions are distributed under the same license as this original. Full details may be found by visiting <https://creativecommons.org/licenses/by-sa/4.0/> or sending a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Acknowledgements

This text began as my sabbatical project in the winter semester of 2012, during which I wrote most of the material for the first four chapters. For the sabbatical leave, I am indebted to Grand Valley State University for its support of the project, as well as to my colleagues in the Department of Mathematics and the College of Liberal Arts and Sciences for their endorsement of the project. I'm also grateful to the American Institute of Mathematics for their support of free and open textbooks in general, and their support of this one in particular.

The beautiful full-color .eps graphics in the text are only possible because of David Austin of GVSU and Bill Casselman of the University of British Columbia. Building on their long-standing efforts to develop tools for high quality mathematical graphics, David wrote a library of Python routines that employ Bill's PiScript program; David's routines are so easy to use that even I could generate graphics like the professionals that he and Bill are. I am deeply grateful to them both.

The current .html version of the text is possible only because of the amazing work of Rob Beezer and his development of the original Mathbook XML, now known as PreTeXt. My ability to take advantage of Rob's work is due to the support of the American Institute of Mathematics, which funded me to attend a weeklong workshop in Mathbook XML in San Jose, CA, in April 2016, as well as the support of the ongoing user group. David Farmer's conversion script saved me hundreds of hours of work by taking my original L^AT_EX source and converting it to PreTeXt; David remains a major source of ongoing support and advocacy. Alex Jordan of Portland Community College has also been a tremendous help, and it is through Alex's fantastic work that live WeBWorK exercises are not only possible, but also included from the 2017 version forward. Mitch Keller of Morningside College agreed in early 2018 to serve as the book's production editor; his technical expertise has contributed to many aspects of the book, including the presence of answers to activities and non- WeBWorK exercises and other supporting material for instructors.

For the 2018 edition, Kathy Yoshiwara of the AIM Editorial Board read the entire text and contributed editorial suggestions for every section. In short, she made the prose cleaner, more direct, and simply better. I'm deeply thankful for her time, effort, and insights.

Over my more than 20 years at GVSU, many of my colleagues have shared with me ideas and resources for teaching calculus. I am particularly indebted to David Austin, Will Dickinson, Paul Fishback, Jon Hodge, and Steve Schlicker for their contributions that improved my teaching of and thinking about calculus, including materials that I have modified and used over many different semesters with students. Parts of these ideas can be found throughout this text. In addition, Will Dickinson and Steve Schlicker provided me access to a large num-

ber of their electronic notes and activities from teaching of differential and integral calculus, and those ideas and materials have similarly impacted my work and writing in positive ways, with some of their problems and approaches finding parallel presentation here.

In the summer of 2012, David and Steve each agreed to write a chapter to support the completion of the material on integral calculus. David is the lead author of Chapter 7 and Steve the lead author of Chapter 8. Along with our colleague Ted Sundstrom, Steve has also contributed a large number of problem and activity solutions and answers. I'm deeply grateful for how the work of these friends and colleagues has made the text so much better.

Shelly Smith of GVSU and Matt Delong of Marian University both provided extensive comments on the first few chapters of early drafts, feedback that was immensely helpful in improving the text. As more and more people use the text, I am grateful to everyone who reads, edits, and uses this book, and hence contributes to its improvement through ongoing discussion.

Finally, I am grateful for all that my students have taught me over the years. Their responses and feedback have helped to shape me as a teacher, and I appreciate their willingness to wholeheartedly engage in the activities and approaches I've tried in class, to let me know how those affect their learning, and to help me learn and grow as an instructor. Early on, they provided useful editorial feedback on this text.

Any and all remaining errors or inconsistencies are mine. I will gladly take reader and user feedback to correct them, along with other suggestions to improve the text.

Contributors

A large and growing number of people have generously contributed to the development or improvement of the text. Contributing authors David Austin and Steven Schlicker have each written drafts of at least one full chapter of the text. Production editor Mitchel Keller has been an indispensable source of technological support and editorial counsel.

The following contributing editors have offered feedback that includes information about typographical errors or suggestions to improve the exposition.

DAVID AUSTIN

GVSU

DAVE KUNG

St. Mary's College of Maryland

RENE ARDILA

GVSU

PAUL LATIOLAIS

Portland State University

ALLAN BICKLE

GVSU

HUGH McGUIRE

GVSU

DAVID CLARK

GVSU

RAY ROSENTRATER

Westmont College

WILL DICKINSON

GVSU

LUIS SANJUAN

*Conservatorio Profesional de Musica de Avila
Spain*

CHARLES FORTIN

*Champlain Regional College
St-Lambert, Quebec, Canada*

STEVEN SCHLICKER

GVSU

MARCI A FROBISH

GVSU

MICHAEL SHULMAN

University of San Diego

PATTI HUNTER

Westmont College

BRIAN STANLEY

Foothill Community College

MITCHEL KELLER

Morningside College

AMY STONE
GVSU

SUE VAN HATTUM
Contra Costa College

ROBERT TALBERT
GVSU

KATHY YOSHIWARA
AIM Editorial Board

GREG THULL
GVSU

Active Calculus: Our Goals

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. As a body of knowledge, calculus has been completely understood 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 their own creative perspective on the subject, it is hardly the case that many of the ideas they present 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 and the age of the internet, 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, a primary purpose of this work is to present a 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*.

Features of the Text

Instructors and students alike will find several consistent features in the presentation, including:

Motivating Questions At the start of each section, we list 2–3 *motivating questions* that provide motivation for why the following material is of interest to us. One goal of each section is to answer each of the motivating questions.

Graphics As much as possible, we strive to demonstrate key fundamental ideas visually, and to encourage students to do the same. Throughout the text, we use full-color¹ graphics to exemplify and magnify key ideas, and to use this graphical perspective alongside both numerical and algebraic representations of calculus.

Summary of Key Ideas Each section concludes with a summary of the key ideas encountered in the preceding section; this summary normally reflects responses to the motivating questions that began the section.

¹To keep cost low, the graphics in the print-on-demand version are in black and white. When the text itself refers to color in images, one needs to view the .html or .pdf electronically.

How to Use this Text

Because the text is free, any professor or student may use the electronic version of the text for no charge.

This text may be used as a stand-alone textbook for a standard first year college calculus sequence. Chapters 1–4 address the standard topics for differential calculus, while Chapters 5–8 correspond to integral calculus, including chapters on differential equations (Chapter 7) and infinite series (Chapter 8).

Contents

Acknowledgements	v
Contributors	vii
Active Calculus: Our Goals	ix
Features of the Text	xi
How to Use this Text	xiii
Functions	iii
0.1 The Definition of Function	iii
0.2 Functions Defined by Tables	iv
0.3 Functions Defined by Graphs	vi
0.4 Functions Defined by Equations	viii
0.5 Function Notation	ix
0.6 Evaluating a Function	x
0.7 Linear Functions	xi
0.8 Describing Functions	xiv
0.9 Function Transformations	xv
0.10 Inverse Functions	xvii
0.11 Exercises	xx
Exponential and Logarithmic Functions	xxi
0.12 Exponential Functions	xxi
0.13 Logarithmic Functions	xxiii

0.14 Exercises	xxv
Trigonometric Functions xxvii	
0.15 Periodic Functions	xxvii
0.16 The Common Trigonometric Functions	xxxi
0.17 Inverse Trigonometric Functions	xxxiii
0.18 Generalized Trigonometric Functions	xxxiv
0.19 Exercises	xxxv
1 Understanding the Derivative 1	
1.1 Introduction to Continuity	1
1.2 Introduction to Limits	9
1.3 How do we measure velocity?	25
1.4 The derivative of a function at a point	41
1.5 The derivative function	64
1.6 Interpreting, estimating, and using the derivative	89
1.7 The second derivative	103
1.8 Differentiability	129
2 Computing Derivatives 139	
2.1 Elementary derivative rules	139
2.2 The product and quotient rules	155
2.3 The sine and cosine functions	172
2.4 Derivatives of other trigonometric functions	184
2.5 The chain rule	195
2.6 Derivatives of Inverse Functions	209
2.7 Derivatives of Functions Given Implicitly	227
2.8 Hyperbolic Functions	244
2.9 The Tangent Line Approximation	249
2.10 The Mean Value Theorem	263
3 Using Derivatives 267	
3.1 Using derivatives to identify extreme values	267
3.2 Global Optimization	289

3.3	Applied Optimization	304
3.4	Using derivatives to describe families of functions	317
3.5	Related Rates	334
3.6	Using Derivatives to Evaluate Limits	355
3.7	Parametric Equations	373
4	The Definite Integral	385
4.1	Determining distance traveled from velocity	385
4.2	Riemann Sums	407
4.3	The Definite Integral	430
4.4	The Fundamental Theorem of Calculus	457
5	Evaluating Integrals	473
5.1	Constructing Accurate Graphs of Antiderivatives	473
5.2	Antiderivatives from Formulas	492
5.3	Differential Equations	505
5.4	The Second Fundamental Theorem of Calculus	511
5.5	Integration by Substitution	528
5.6	Integration by Parts	545
5.7	Other Options for Finding Algebraic Antiderivatives	564
5.8	Numerical Integration	578
6	Using Definite Integrals	599
6.1	Using Definite Integrals to Find Area and Length	599
6.2	Using Definite Integrals to Find Volume	620
6.3	Density, Mass, and Center of Mass	640
6.4	Physics Applications: Work, Force, and Pressure	658
6.5	Improper Integrals	677
7	Differential Equations	695
7.1	An Introduction to Differential Equations	695
7.2	Qualitative behavior of solutions to DEs	709
7.3	Euler's method	729
7.4	Separable differential equations	748

7.5	Modeling with differential equations	765
7.6	Population Growth and the Logistic Equation	781
8	Sequences and Series	797
8.1	Sequences	797
8.2	Geometric Series	812
8.3	Series of Real Numbers	830
8.4	Alternating Series	859
8.5	Taylor Polynomials and Taylor Series	882
8.6	Power Series	908
A	A Short Table of Integrals	929
Index		933

The material in this section represents material that is covered in a typical precalculus course. For a more extensive treatment of PreCalculus we refer the reader to PreCalculus at Nebraska.

Functions

0.1 The Definition of Function

We often want to predict values of one variable from the values of a related variable. For a more extensive treatment of functions we refer the reader to PreCalculus at Nebraska: Introduction to Functions

Definition of Function

A **function** is a relationship between two variables for which a unique value of the **output** variable can be determined from a value of the **input** variable.

What distinguishes functions from other variable relationships? The definition of a function calls for a *unique value* that is, *exactly one value* of the output variable corresponding to each value of the input variable. This property makes functions useful in applications because they can often be used to make predictions.

A function can be described in several different ways. In the following sections, we consider examples of functions defined by tables, by graphs, and by equations.

0.2 Functions Defined by Tables

When we use a table to describe a function, the first variable in the table (the left column of a vertical table or the top row of a horizontal table) is the input variable, and the second variable is the output. We say that the output variable *is a function of* the input.

Example 0.2.1.

- a Table 0.2.2 shows data on sales compiled over several years by the accounting office for Eau Claire Auto Parts, a division of Major Motors. In this example, the year is the input variable, and total sales is the output. We say that total sales, S , *is a function of* t .

Year (t)	Total sales (S)
2000	\$612,000
2001	\$663,000
2002	\$692,000
2003	\$749,000
2004	\$904,000

Table 0.2.2

- b Table 0.2.3 gives the cost of sending printed material by first-class mail in 2016.

Weight in ounces (w)	Postage (P)
$0 < w \leq 1$	\$0.47
$1 < w \leq 2$	\$0.68
$2 < w \leq 3$	\$0.89
$3 < w \leq 4$	\$1.10
$4 < w \leq 5$	\$1.31
$5 < w \leq 6$	\$1.52
$6 < w \leq 7$	\$1.73

Table 0.2.3

If we know the weight of the article being shipped, we can determine the required postage from Table 0.2.3. For instance, a catalog weighing 4.5 ounces would require \$1.31 in postage. In this example, w is the input variable and p is the output variable. We say that p *is a function of* w .

- c Table 0.2.4 records the age and cholesterol count for 20 patients tested in a hospital survey.

Age	Cholesterol count	Age	Cholesterol count
53	217	51	209
48	232	53	241
55	198	49	186
56	238	51	216
51	227	57	208
52	264	52	248
53	195	50	214
47	203	56	271
48	212	53	193
50	234	48	172

Table 0.2.4

According to these data, cholesterol count is *not* a function of age, because several patients who are the same age have different cholesterol levels. For example, three different patients are 51 years old but have cholesterol counts of 227, 209, and 216, respectively. Thus, we cannot determine a *unique* value of the output variable (cholesterol count) from the value of the input variable (age). Other factors besides age must influence a persons cholesterol count.

0.3 Functions Defined by Graphs

A graph may also be used to define one variable as a function of another. The input variable is displayed on the horizontal axis, and the output variable on the vertical axis.

Example 0.3.1. Figure 0.3.2 shows the number of hours, H , that the sun is above the horizon in Peoria, Illinois, on day t , where January 1 corresponds to $t = 0$.

- a The input variable, t , appears on the horizontal axis. The number of daylight hours, H , is a function of the date. The output variable appears on the vertical axis.
- b The point on the curve where $t = 150$ has $H \approx 14.1$, so Peoria gets about 14.1 hours of daylight when $t = 150$, which is at the end of May.
- c $H = 12$ at the two points where $t \approx 85$ (in late March) and $t \approx 270$ (late September).
- d The maximum value of 14.4 hours occurs on the longest day of the year, when $t \approx 170$, about three weeks into June. The minimum of 9.6 hours occurs on the shortest day, when $t \approx 355$, about three weeks into December.

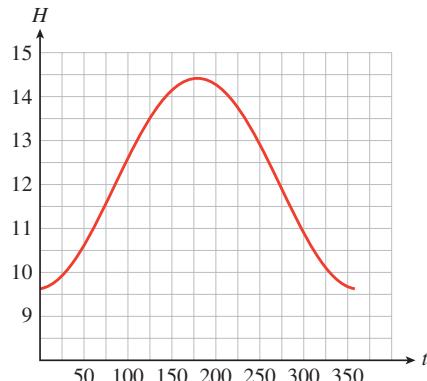


Figure 0.3.2

We have a method of quickly determining if a relationship is a function once we have a graph of the relationship.

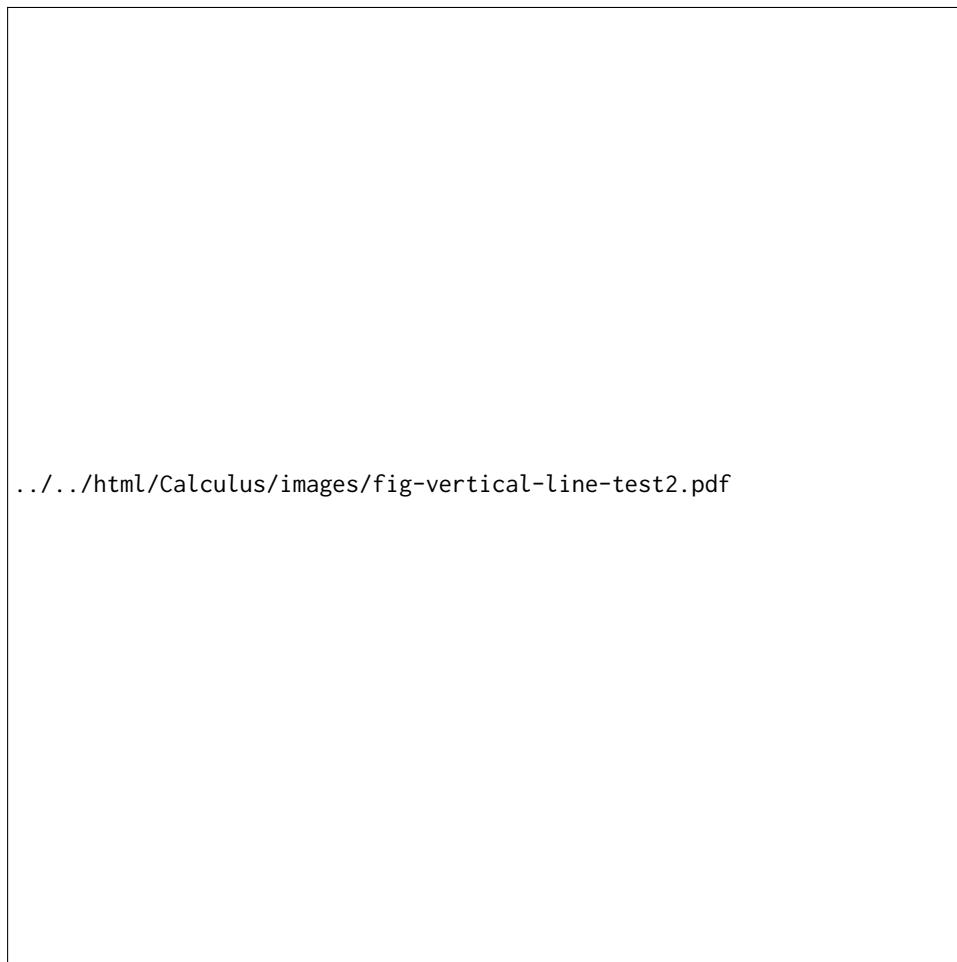
The Vertical Line Test

A graph represents a function if and only if every vertical line intersects the graph in at most one point.

Geogebra: www.geogebra.org/m/j9j59n7z

Figure 0.3.3: Demonstration of the Vertical Line Test

Example 0.3.4. Use the vertical line test to decide which of the graphs in Figure 0.3.5 represent functions.

**Figure 0.3.5****Solution.**

- Graph (a) represents a function, because it passes the vertical line test.
- Graph (b) is not the graph of a function, because the vertical line at (for example) $x = 1$ intersects the graph at two points.
- For graph (c), notice the break in the curve at $x = 2$: The solid dot at $(2, 1)$ is the only point on the graph with $x = 2$; the open circle at $(2, 3)$ indicates that $(2, 3)$ is not a point on the graph. Thus, graph (c) is a function, with $f(2) = 1$.

0.4 Functions Defined by Equations

Example 0.4.1 illustrates a function defined by an equation.

Example 0.4.1. As of 2016, One World Trade Center in New York City is the nations tallest building, at 1776 feet. If an algebra book is dropped from the top of One World Trade Center, its height above the ground after t seconds is given by the equation

$$h = 1776 - 16t^2$$

Thus, after 1 second the books height is

$$h = 1776 - 16(1)^2 = 1760 \text{ feet}$$

After 2 seconds its height is

$$h = 1776 - 16(2)^2 = 1712 \text{ feet}$$

For this function, t is the input variable and h is the output variable. For any value of t , a unique value of h can be determined from the equation for h . We say that h is a *function of t* .

0.5 Function Notation

There is a convenient notation for discussing functions. First, we choose a letter, such as f , g , or h (or F , G , or H), to name a particular function. (We can use any letter, but these are the most common choices.) For instance, the height, h , of a falling textbook is a function of the elapsed time, t . We might call this function f . In other words, f is the name of the relationship between the variables h and t . We write

$$h = f(t)$$

which means " h is a function of t , and f is the name of the function."

Example 0.5.1. In Example 0.4.1, the height of an algebra book dropped from the top of One World Trade Center is given by the equation

$$h = 1776 - 16t^2$$

We see that

$$\begin{array}{ll} \text{when } t = 1 & h = 1760 \\ \text{when } t = 2 & h = 1712 \end{array}$$

Using function notation, these relationships can be expressed more concisely as

$$f(1) = 1760 \quad \text{and} \quad f(2) = 1712$$

which we read as " f of 1 equals 1760" and " f of 2 equals 1712." The values for the input variable, t , appear *inside* the parentheses, and the values for the output variable, h , appear on the other side of the equation.

Remember that when we write $y = f(x)$, the symbol $f(x)$ is just another name for the output variable.

0.6 Evaluating a Function

Finding the value of the output variable that corresponds to a particular value of the input variable is called **evaluating the function**.

Example 0.6.1. Let g be the name of the postage function defined by Table 0.2.3. Find $g(1)$, $g(3)$, and $g(6.75)$.

Solution. According to the table,

$$\begin{array}{lll} \text{when } w = 1, & p = 0.47 & \text{so } g(1) = 0.47 \\ \text{when } w = 3, & p = 0.89 & \text{so } g(3) = 0.89 \\ \text{when } w = 6.75, & p = 1.73 & \text{so } g(6.75) = 1.73 \end{array}$$

Thus, a letter weighing 1 ounce costs \$0.47 to mail, a letter weighing 3 ounces costs \$0.89, and a letter weighing 6.75 ounces costs \$1.73.

If a function is described by an equation, we simply substitute the given input value into the equation to find the corresponding output, or function value.

Example 0.6.2. The function H is defined by $H = f(s) = \frac{\sqrt{s+3}}{s}$. Evaluate the function at the following values.

a $s = 6$

b $s = -1$

Solution.

a $f(6) = \frac{\sqrt{6+3}}{6} = \frac{\sqrt{9}}{6} = \frac{3}{6} = \frac{1}{2}$. Thus, $f(6) = \frac{1}{2}$.

b $f(-1) = \frac{\sqrt{-1+3}}{-1} = \frac{\sqrt{2}}{-1} = -\sqrt{2}$. Thus, $f(-1) = -\sqrt{2}$.

To simplify the notation, we sometimes use the same letter for the output variable and for the name of the function.

0.7 Linear Functions

Linear relationships are relationships in which the rate of change is constant.

Linear Equation

A **linear function** is a function which has a **constant rate of change**.

Many phenomena can be modeled using linear functions $y = f(x)$ where the equations have the form

$$f(x) = (\text{starting value}) + (\text{rate of change}) \cdot x.$$

The initial value, or the value of $f(0)$, is the vertical intercept of the graph, and the rate of change is the slope of the graph. Thus, we can write an equation of a line as

$$f(x) = b + mx$$

where the constant term, b , is the vertical intercept of the line, and m , the coefficient of x , is the slope of the line. This form for an equation of a line is called the **slope-intercept form**.

Slope-Intercept Form

If we write an equation of a linear function in the form,

$$f(x) = b + mx$$

then m is the **slope** of the line, and b is the **vertical intercept**.

Example 0.7.1. An icicle grows according to the formula $H(t) = 0.05t + 0.12$, where t is the time in minutes since the first measurement was taken, and $H(t)$ is the height of the icicle in centimeters.

- a The slope is 0.05, which tells us that the icicle's height grows by 0.05 cm each minute.
- b The y -intercept is 0.12, which tells us that the height of the icicle was 0.12 cm at the first measurement.

Example 0.7.2. Samantha owns a catering business. For any party with up to 100 guests, she charges 2,000. She charges 8 per person for each additional guest over 100. Give a formula for the cost of having Samantha cater your party as a function of the number of guests over 200.

Solution. A possible formula for the cost of having Samantha cater your party is $C(g) = 2000 + 8g$, where g is the number of guests over 100.

The following formula provides a method to finding the value of the slope m when given two points on the line.

Two-Point Slope Formula

The slope of the line passing through the points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is given by

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}, \quad x_2 \neq x_1.$$

Example 0.7.3. Compute the slope of the line between the points $(6, -3)$ and $(4, 3)$.

Solution. Substitute the coordinates of Q_1 and Q_2 into the slope formula to find

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{3 - (-3)}{4 - 6} = \frac{6}{-2} = -3.$$

This value for the slope, -3 , is the same value found above.

Sometimes you will be given the slope of a line and another point on that line. The following formula is helpful in that situation.

Point-Slope Form

An equation of the line that passes through the point (x_1, y_1) and has slope m is

$$y = y_1 + m(x - x_1).$$

Example 0.7.4. Find a formula for the line that has slope -2 and passes through the point $(-1, 4)$

Solution. Using point-slope form, we have the line $y = 4 + (-2)(x - (-1))$.

We can also write this in slope-intercept form by simplifying: $y = 2 - 2x$.

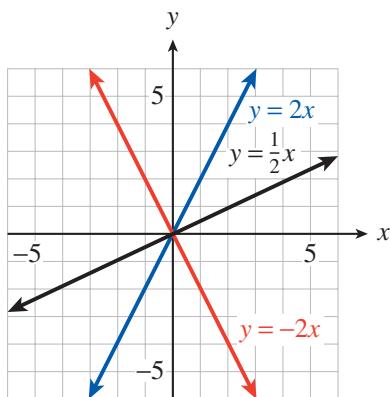
It is also useful to introduce the term x -intercept.

 x -intercept

An **x -intercept** for a function $f(x)$ is the value of x such that $f(x) = 0$.

Note 0.7.5. In the equation $f(x) = b + mx$, we call m and b **parameters**. Their values are fixed for any particular linear equation; for example, in the equation $y = 2x + 3$, $m = 2$ and $b = 3$, and the variables are x and y . By changing the values of m and b , we can write the equation for any line except a vertical line (see Figure 0.7.6). The collection of all linear functions $f(x) = b + mx$ is called a **two-parameter** family of functions.

These lines have the same y-intercept but different slopes.



These lines have the same slope but different y-intercepts.

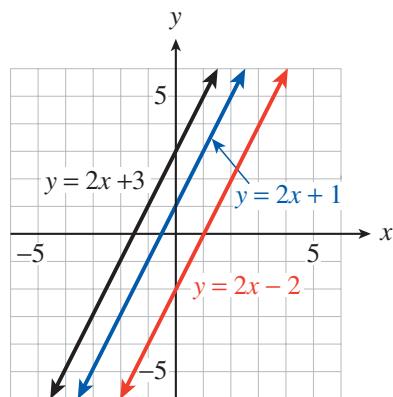


Figure 0.7.6

0.8 Describing Functions

There are several terms that will be useful in describing functions. We first begin with the notion of an increasing function.

Increasing Function

A function f is increasing if the values of $f(x)$ increase as x increases. The graph of an increasing function climbs as we move from left to right.

Decreasing Function

A function f is decreasing if the values of $f(x)$ decrease as x increases. The graph of a decreasing function falls as we move from left to right.

Monotonic Function

A function $f(x)$ is monotonic if it increases for all x or decreases for all x .

Directly Proportional

We say y is directly proportional to x if there is a nonzero constant k such that, $y = kx$. This k is called the constant of proportionality.

Inversely Proportional

We say that y is inversely proportional to x if y is proportional to the reciprocal of x , that is, $y = \frac{k}{x}$ for a nonzero constant k .

0.9 Function Transformations

It is also useful to talk about transformations of functions. Several key facts will be useful.

Function Transformations

1. Multiplying a function by a constant, c , stretches the graph vertically (if $c > 1$) or shrinks the graph vertically (if $0 < c < 1$)
2. A negative sign (if $c < 0$) reflects the graph about the x -axis, in addition to shrinking or stretching.
3. Replacing y by $(y + k)$ moves a graph up by k (down if k is negative).
4. Replacing x by $(x - h)$ moves a graph to the right by h (to the left if h is negative).

Example 0.9.1. Let $f(x) = x^2$. We will explore different transformations performed on the graph of $f(x)$.

- $g(x) = 2(x) = 2x^2$ is a vertical stretch of $f(x)$ by a factor of 2.
- $h(x) = -2(x) = -2x^2$ is a vertical stretch of $f(x)$ by a factor of 2 and a reflection across the x -axis.
- $j(x) = f(x) + 5 = x^2 + 5$ is a vertical shift of $f(x)$ up 5 units.
- $k(x) = f(x + 5) = (x + 5)^2$ is a horizontal shift of $f(x)$ left 5 units.

Order of Transformations

Suppose that $f(x)$ is our function we are applying transformations to. Once written in the form

$$a \cdot f(b \cdot (x + h)) + k$$

the order of transformations is:

1. horizontal stretch or compress by a factor of b or $\frac{1}{b}$ (if $b < 0$ then also reflect about y -axis)
2. shift horizontally left/ right by h
3. vertically stretch or compress by a factor of a or $\frac{1}{a}$ (if $a < 0$ then also reflect about x -axis)
4. shift vertically up/ down by k

Example 0.9.2. Describe the function

$$5 \cdot [f(-3x - 6) - 1]$$

as a list of transformations done to $f(x)$ in the appropriate order.

Solution. First, let's rewrite the function in the form given above:

$$\begin{aligned} 5 \cdot [f(-3x - 6) - 1] &= 5f(-3x - 6) - 5 \text{ we distributed in the 5} \\ &= 5f(-3(x + 2)) - 5 \text{ we factored out the -3} \end{aligned}$$

Now that the function is written in the desired order, we may list off the transformations in the correct order using the order discussed above:

1. horizontal compress by 3 and reflect about the y -axis
2. shift horizontally left by 2
3. vertically stretch by 5
4. shift vertically down by 5

0.10 Inverse Functions

It is also useful to consider inverse functions.

Inverse Functions

Suppose the inverse of f is a function, denoted by f^{-1} . Then

$$f^{-1}(y) = x \text{ if and only if } f(x) = y.$$

Example 0.10.1. Suppose g is the inverse function for f , and we know the following function values for f :

$$f(-3) = 5, \quad f(2) = 1, \quad f(5) = 0.$$

Find $g(5)$ and $g(0)$.

Solution. We know that $g(5) = -3$ because $f(-3) = 5$, and $g(0) = 5$ because $f(5) = 0$. Tables may be helpful in visualizing the two functions, as shown below.

$y = f(x)$	
x	y
-3	5
2	1
5	0

→ Interchange the columns →

$x = g(y)$	
y	x
5	-3
1	2
0	5

For the function f , the input variable is x and the output variable is y . For the inverse function g , the roles of the variables are interchanged: y is now the input and x is the output.

Functions and Inverse Functions

Suppose f^{-1} is the inverse function for f . Then

$$f^{-1}(f(x)) = x \text{ and } f(f^{-1}(y)) = y$$

as long as x is in the domain of f , and y is in the domain of f^{-1} .

We also have a method of quickly determining if a function is invertible once we have a graph of the function.

Horizontal Line Test

If no horizontal line intersects the graph of a function more than once, then its inverse is also a function.

Figure 0.10.2: Notice that as the line moves up the y -axis, it only ever intersects the graph in a single place. This means this function is invertible.

Figure 0.10.3: Notice that as the line moves up the y -axis, it sometimes intersects the graph in more than one place. This means this function is not invertible.

Example 0.10.4. Which of the functions in Figure 0.10.5 are invertible?

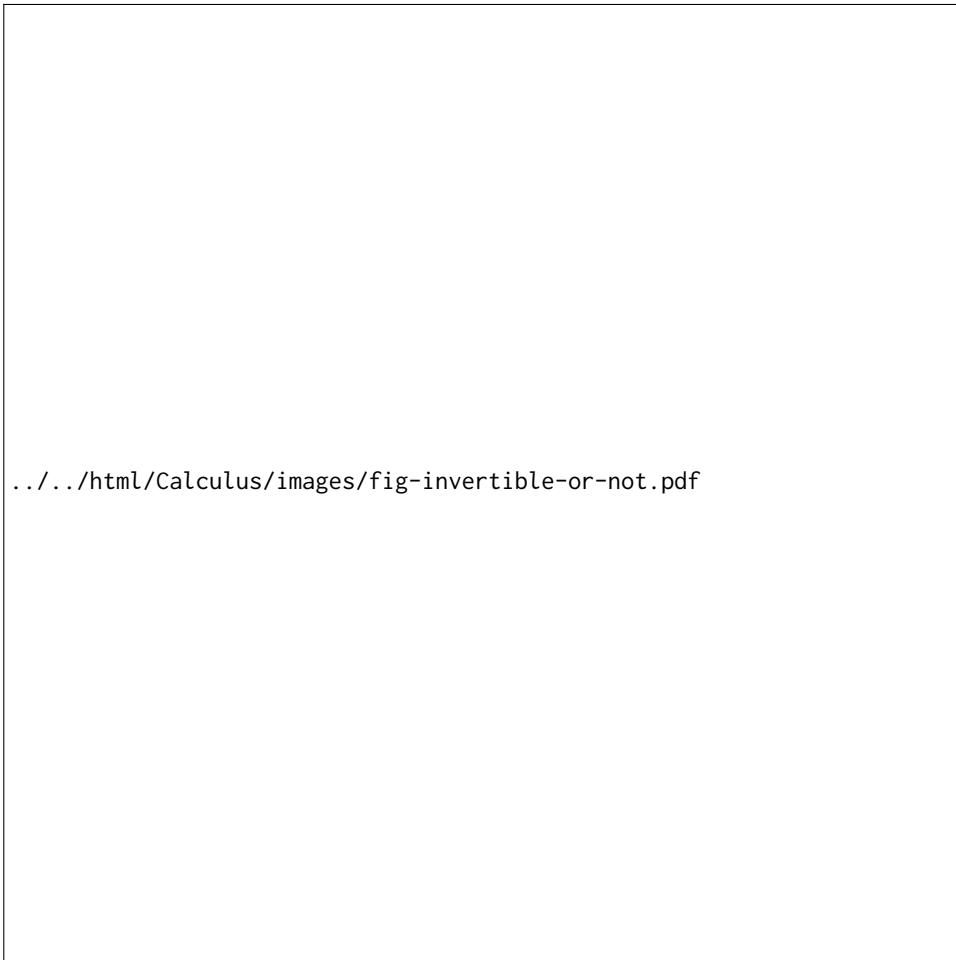


Figure 0.10.5

Solution. In each case, apply the horizontal line test to determine whether the function is invertible. Because no horizontal line intersects their graphs more than once, the functions pictured in Figures 0.10.5(a) and (c) are invertible. The functions in Figures 0.10.5(b) and (d) are not invertible.

We have been talking about how to tell if the inverse of a function is also a function, but in practice this is not the language typically used. Usually we ask this same question in the form "Is the function invertible?" The following definition explains this relationship:

Invertible Function

If $y = f(x)$ is a function such that its inverse, $x = f^{-1}(y)$, is also a function then we say that $f(x)$ is an **invertible function**.

The inverse function f^{-1} undoes the effect of the function f . The function $f(t) = 6 + 2t$ multiplies the input by 2 and then adds 6 to the result. The inverse function $f^{-1}(H) = \frac{H - 6}{2}$ undoes those operations in reverse order: It subtracts 6 from the input and then divides the result by 2.

If we apply the function f to a given input value and then apply the function f^{-1} to the output from f , the end result will be the original input value. For example, if we choose $t = 5$ as an input value, we find that

$$f(5) = 6 + 2(5) = 16$$

Multiply by 2, then add 6.

$$\text{and } f^{-1}(16) = \frac{16 - 6}{2} = 5.$$

Subtract 6, then divide by 2.

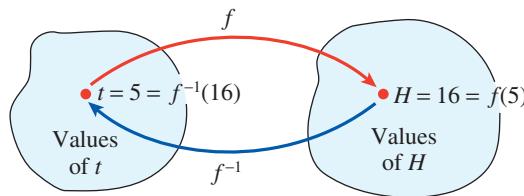


Figure 0.10.6

We return to the original input value, 5.

0.11 Exercises

1. (Slope and Intercept) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Graphs of Linear Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Proportionality) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Finding Lines) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Transformations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Matching Transformations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

Exponential and Logarithmic Functions

In this chapter we define exponential and logarithmic functions. For a more extensive treatment of exponential functions we refer the reader to PreCalculus at Nebraska: Exponential Functions and for a more extensive treatment of exponential functions we refer the reader to PreCalculus at Nebraska: Logarithmic Functions

0.12 Exponential Functions

Before we discuss exponential functions, it is first useful to define the notion of concavity.

Concavity

The graph of a function is concave up if it bends upward as we move left to right; it is concave down if it bends downward. (See the figure for four possible shapes.) A line is neither concave up nor concave down.

An exponential function will be either concave up everywhere! We define such an exponential as follows:

Exponential Function

$$P(t) = a \cdot b^t, \text{ where } b > 0 \text{ and } b \neq 1, a \neq 0.$$

The constant a is the y -value of the y -intercept of the function.

The equation

$$P(t) = P_0 a^t$$

gives an exponential function with base a . Then

$$\frac{P(t+1)}{P(t)} = \frac{P_0 a^{t+1}}{P_0 a^t} = a = \text{constant rate of growth/decay}$$

- Growth: $a > 1$: Doubling time: the time it takes to double the initial amount

- Decay $a < 1$: Half-life: the time it takes to decay to half of the initial amount

Doubling Time and Half-life

The doubling time of an exponentially increasing quantity is the time required for the quantity to double. The half-life of an exponentially decaying quantity is the time required for the quantity to be reduced by a factor of one half.

We will (naturally) consider exponentials (and later, logarithms) in base

$$e \approx 2.718281828459\dots \text{ (irrational number)}$$

For example, the general exponential function in the natural base

$$P(t) = P_0 a^t = P_0 (e^k)^t = P_0 e^{kt}, \quad a = e^k.$$

Thus, we will have

- exponential growth if $a > 1$, which gives $k > 0$
- exponential decay if $a < 1$, which gives $k < 0$

The number k is called *the continuous rate of growth/decay*. To find k we will need the logarithmic function.

0.13 Logarithmic Functions

Recall inverse functions. If for each y in the range of f there exists exactly one value of x such that $f(x) = y$, then f has an inverse at y denoted by f^{-1} such that

$$f(x) = y \iff f^{-1}(y) = x$$

Hence

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

The *inverse* of the exponential function $f(x) = e^x$ is the natural *logarithmic* function $f^{-1}(x) = \ln(x)$ so we have

$$\ln(x) = c \iff e^c = x$$

The Logarithm

Let $b \neq 1$ be a positive number, then the function

$$f(t) = \log_b(t)$$

is called a logarithm with base b .

Upon inputting a value t the function $\log_b(t)$ will tell you the power of b which will yield t .

Due to the relationship between logarithms and exponentials, we often say that the equations

$$x = \log_b(y) \quad \text{and} \quad b^x = y$$

are equivalent.

Warning 0.13.1. Two common bases of a logarithm are base 10 and e . Since they are used so often, we have developed a short hand notation for a logarithm of base 10 and a logarithm of base e . This short hand is shown below:

$$\log_{10}(y) = \log(y).$$

$$\log_e(y) = \ln(y).$$

In other words, we simply drop the subscript when referring to base 10 and we change to \ln when referring to base e

There are some important properties of logarithms that you should be familiar with.

Properties of Logarithms

If $x, y, b > 0$, and $b \neq 1$, then

$$1 \quad \log_b(xy) = \log_b(x) + \log_b(y),$$

$$2 \quad \log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y),$$

$$3 \quad \log_b(x^k) = k \cdot \log_b(x),$$

$$4 \log_b(b^y) = y,$$

$$5 \ b^{\log_b(x)} = x.$$

0.14 Exercises

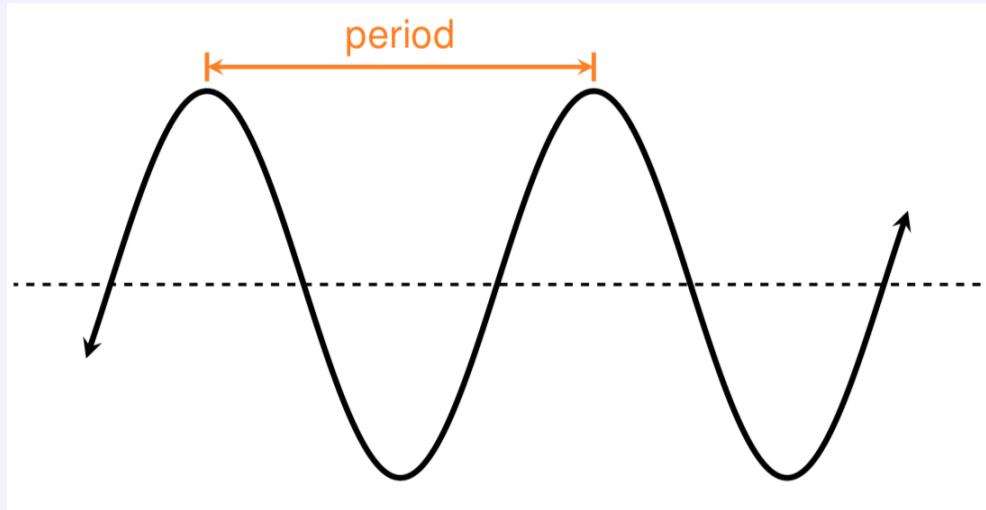
1. (General Exponential Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Finding Exponential Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Finding the Parameters for an Exponential Function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Half-Life) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Applied Half-Life) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Solving Exponential Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Using Inverse Functions With Exponentials and Logarithms) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

Trigonometric Functions

In this chapter we define trigonometric functions. For a more extensive treatment of trigonometric functions we refer the reader to PreCalculus at Nebraska: College Trigonometry.

0.15 Periodic Functions

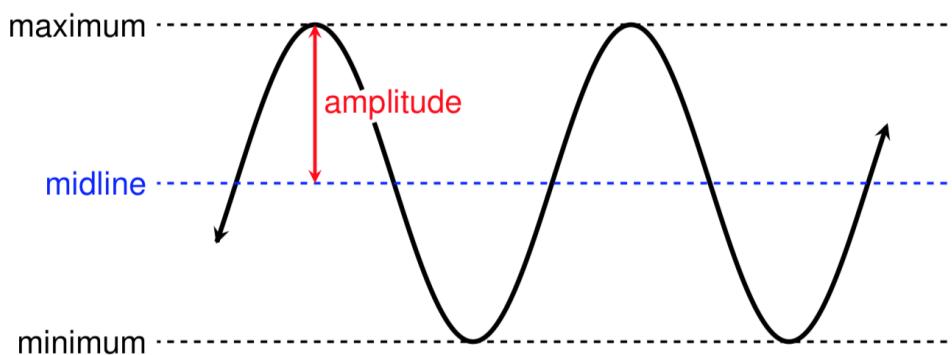
A **periodic function** is a function whose output values repeat on a regular interval. In mathematical terms, we say that a periodic function is a function for which a specific horizontal shift, P , results in the original function: $f(x + P) = f(x)$ for all values of x . The **period** of a periodic function is how long it takes for the output values to begin repeating, or the smallest horizontal shift with $P > 0$, such that $f(x + P) = f(x)$ for all values of x .



In addition to the period we also define the midline and amplitude of a periodic function.

The **midline** of a periodic function is the horizontal line halfway, or midway, between the function's maximum and minimum output values.

The **amplitude** of a periodic function is the distance between the function's maximum (or minimum) output value and the midline.

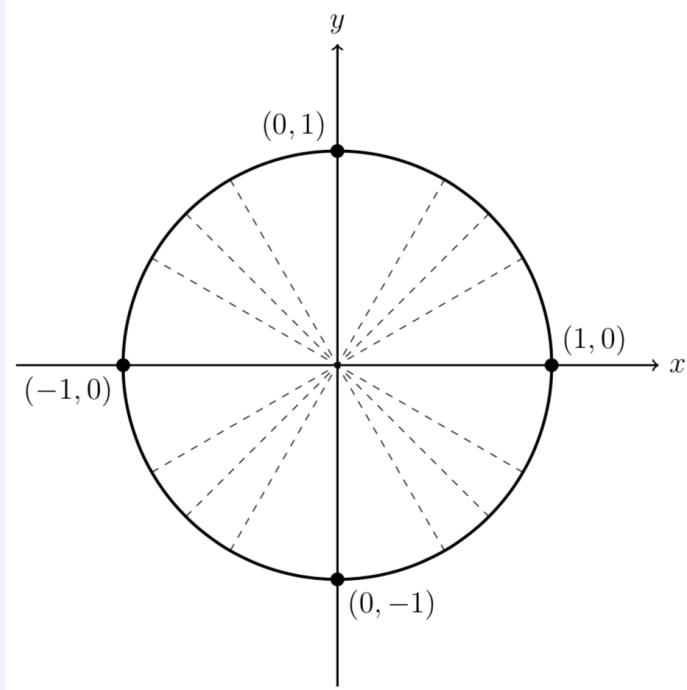


The midline often represents the average value of a wave-like periodic function and should always be written as the equation of a horizontal line. The amplitude of a periodic function represents how "large" the function's oscillations are.

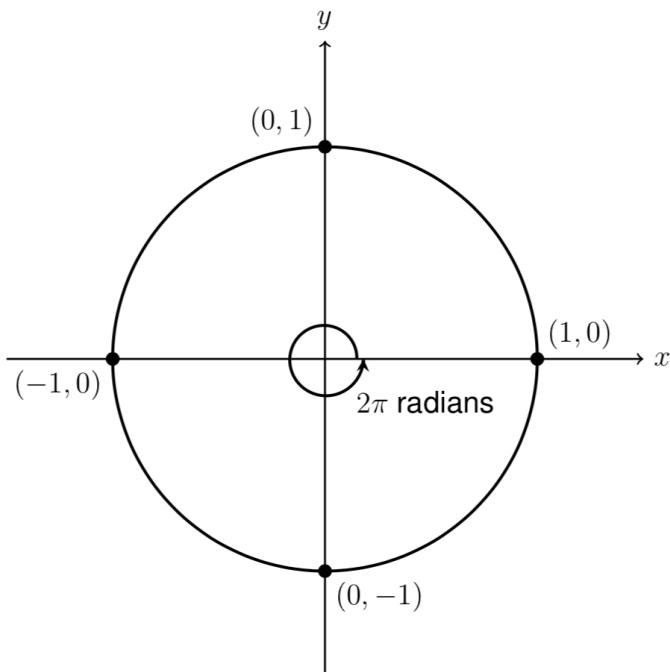
While measuring angles in degrees may be familiar to many students, doing so often complicates matters since the units of measure can get in the way of calculations. For this reason, another measure of angles is commonly used. This measure is based on the distance around the unit circle.

The Unit Circle

The **unit circle** is a circle of radius 1, centered at the origin of the xy -plane. When measuring an angle around the unit circle, we travel in the counterclockwise direction, starting from the positive x -axis. A negative angle is measured in the opposite, or clockwise, direction. A complete trip around the unit circle amounts to a total of 360 degrees.



A **radian** is a measurement of an angle that arises from looking at angles as a fraction of the circumference of the unit circle. A complete trip around the unit circle amounts to a total of 2π radians.



Radians are a **unitless measure**. Therefore, it is not necessary to write the label "radians" after a radian measure, and if you see an angle that is not labeled with "degrees" or the degree symbol, you should assume that it is a radian measure.

Radians and degrees both measure angles. Thus, it is possible to convert between the two. Since one rotation around the unit circle equals 360 degrees or 2π radians, we can use this as a conversion factor.

Converting Between Radians and Degrees

Since 360 degrees $= 2\pi$ radians, we can divide each side by 360 and conclude that

$$1 \text{ degree} = \frac{2\pi \text{ radians}}{360} = \frac{\pi \text{ radians}}{180}$$

So, to convert from degrees to radians, we can multiply by $\frac{\pi \text{ radians}}{180^\circ}$

Similarly, we can conclude that

$$1 \text{ radian} = \frac{360^\circ}{2\pi} = \frac{180^\circ}{\pi}$$

So, to convert from radians to degrees, we can multiply by $\frac{180^\circ}{\pi \text{ radians}}$

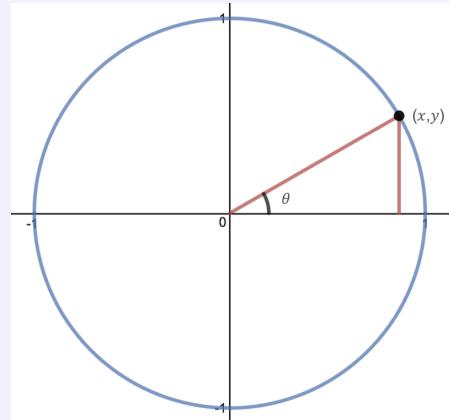
0.16 The Common Trigonometric Functions

The Sine and Cosine Functions

Given an angle θ (in either degrees or radians) and the (x, y) coordinates of the corresponding point on the unit circle, we define cosine and sine as

$$\cos(\theta) = x \quad \text{and} \quad \sin(\theta) = y$$

Note that sine and cosine are **functions** that take angles as inputs.



Using the (x, y) coordinates we can find the cosine and sine values of common angles on the unit circle.

The applet below shows how the graphs of sine and cosine relate to the x and y coordinates of the point P on the unit circle. Move the slider to change the angle θ and observe how the x and y values of point P relate to the outputs of $\cos(\theta)$ and $\sin(\theta)$.

Geogebra: www.geogebra.org/m/pJqv9pR

Figure 0.16.1: How the Graphs of Sine and Cosine Relate to the Unit Circle

In addition to sine and cosine it is also useful to define several additional trigonometric functions.

The Tangent Function

Given an angle θ (in either degrees or radians) and the (x, y) coordinates of the corresponding point on the unit circle, we define tangent as

$$\tan(\theta) = \frac{y}{x}$$

Just like sine and cosine, tangent is a **function** that takes angles as inputs.

.../.../html/Calculus/images/imagesChap13/Sinea

Note that since $\sin(\theta) = y$ and $\cos(\theta) = x$, we can also define tangent as

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$$

We can find the tangent values of common angles on the unit circle by using the sine and cosine values of the angles (or the corresponding y and x coordinates).

There are times when it can be useful to consider functions like $\frac{1}{\sin(\theta)}$ or $\frac{1}{\cos(\theta)}$. This happens often enough that mathematicians have created special names for these types of functions. Generally, these functions are called reciprocal trigonometric functions, and they can be defined in terms of the sine, cosine, and tangent functions.

The Reciprocal Trigonometric Functions

Given an angle θ , we define:

the secant function as

$$\sec(\theta) = \frac{1}{\cos(\theta)}$$

the cosecant function as

$$\csc(\theta) = \frac{1}{\sin(\theta)}$$

the cotangent function as

$$\cot(\theta) = \frac{1}{\tan(\theta)}$$

0.17 Inverse Trigonometric Functions

In the previous subsection, we evaluated trigonometric functions at various angles, but what if we needed to know what angle yields a specific sine, cosine, or tangent value. To find angles, we need inverse trigonometric functions.

The Inverse Sine, Inverse Cosine, and Inverse Tangent Functions

If $\sin(\theta) = a$, then $\sin^{-1}(a) = \theta$ for $-1 \leq a \leq 1$

If $\cos(\theta) = a$, then $\cos^{-1}(a) = \theta$ for $-1 \leq a \leq 1$

If $\tan(\theta) = a$, then $\tan^{-1}(a) = \theta$ for $-\infty < a < \infty$

Note that the *output* of each of these inverse functions is an *angle*.

Remark 0.17.1. Do not confuse $\sin^{-1}(a)$ with $\frac{1}{\sin(a)}$ as they are not equivalent. The same can be said with $\cos^{-1}(a)$ and $\tan^{-1}(a)$. As a result, we introduce an alternative (common) notation for the inverse trigonometric functions.

Note 0.17.2 (Common Notation for the Inverse Trigonometric Functions). The inverse sine function, $\sin^{-1}(a)$ is sometimes called the **arcsine** function, and notated $\arcsin(a)$.

The inverse cosine function, $\cos^{-1}(a)$ is sometimes called the **arccosine** function, and notated $\arccos(a)$.

The inverse tangent function, $\tan^{-1}(a)$ is sometimes called the **arctangent** function, and notated $\arctan(a)$.

0.18 Generalized Trigonometric Functions

Like all functions, trigonometric functions can be transformed by changing properties like the period, midline, and amplitude of the function. In this subsection, we explore transformations of the sine and cosine functions and use them to model real life situations.

Transformations of Sine and Cosine

Given an equation in the form

$$f(t) = A \sin(Bt) + k \quad \text{or} \quad g(t) = A \cos(Bt) + k$$

$|A|$ is the vertical stretch/compression and the **amplitude** of the function.

$|B|$ is the horizontal stretch/compression and is **related** to the **period** of the function, P , by the formula

$$P = \frac{2\pi}{|B|}$$

k is the vertical shift and determines the **midline** of the function.

0.19 Exercises

1. (Period and Amplitude) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Period and Amplitude) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Finding Trigonometric Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Finding Trigonometric Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (The Unit Circle) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Trigonometric Functions as Compositions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

Contents

Chapter 1

Understanding the Derivative

1.1 Introduction to Continuity

Motivating Questions

- What does it mean to say that a function f is continuous at $x = a$?
 - What does it mean to say that a function f is continuous everywhere?
-

In this section we introduce the idea of a continuous function. Many of the results in calculus require that the functions be continuous so having a strong understanding of continuous functions will be very important.

1.1.1 Being continuous at a point

Intuitively, a function is continuous if we can draw its graph 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.

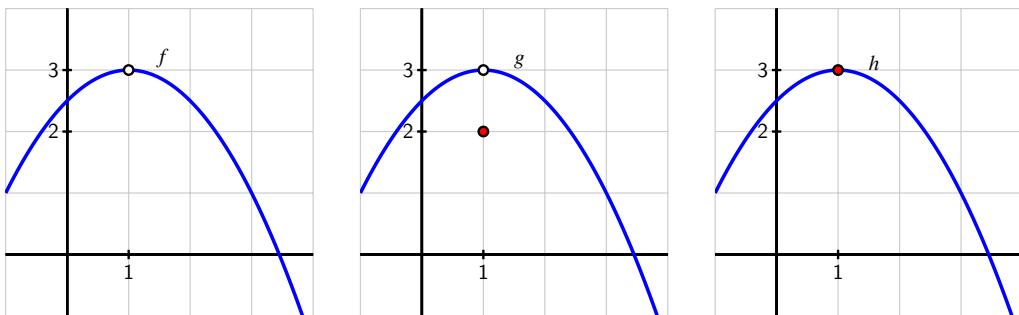


Figure 1.1.1: Functions f , g , and h that demonstrate subtly different behaviors at $a = 1$.

First consider the function in the left-most graph. Note that $f(1)$ is not defined, which leads to the resulting hole in the graph of f at $a = 1$. If you were to draw the graph of f yourself then you would need to lift your pencil when you reached $f(1)$. We will naturally say that f is *not continuous* at $a = 1$. For the function g , we observe that while $g(1)$ is defined, the value of $g(1) = 2$ is not what you would "expect." Specifically, you would expect $g(1)$ to be 3, not 2. Thus, to draw the graph of g you would need to lift your pencil at $a = 1$. Again, we will say that g is *not continuous* at $a = 1$, 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. In this case we would say that h is continuous at $a = 1$.

The above examples demonstrate a discontinuity commonly known as a *removable discontinuity*. This is, however, not the only way in which a function can be discontinuous. Another type of discontinuity is the so called *jump discontinuity*.

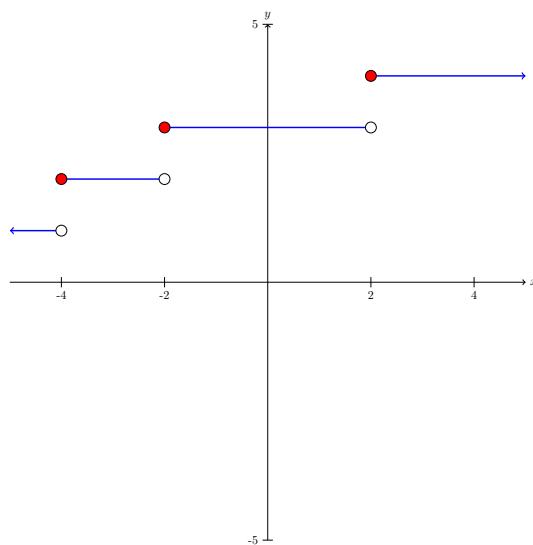


Figure 1.1.2: Places where functions make a "jump" also produce discontinuities.

A third type of discontinuity is the so called *infinite* discontinuity. Infinite discontinuities exist when the points values of a function diverge to infinity. A classic example of an infinite discontinuity is the point $x = 0$ for the function $f(x) = 1/x$.

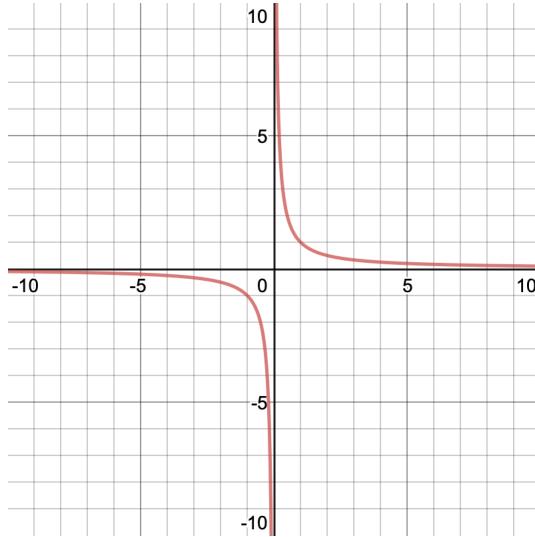


Figure 1.1.3: A plot of the function $y = f(x)$ for the function $f(x) = 1/x$.

Example 1.1.4. Consider the function

$$f(x) = \frac{3x(x+4)}{x^2 - 16}.$$

Are there x -values where the function is discontinuous? If so, what are those x -values and why type of discontinuity occurs at that value?

Hint. Consider the graph of $y = f(x)$.

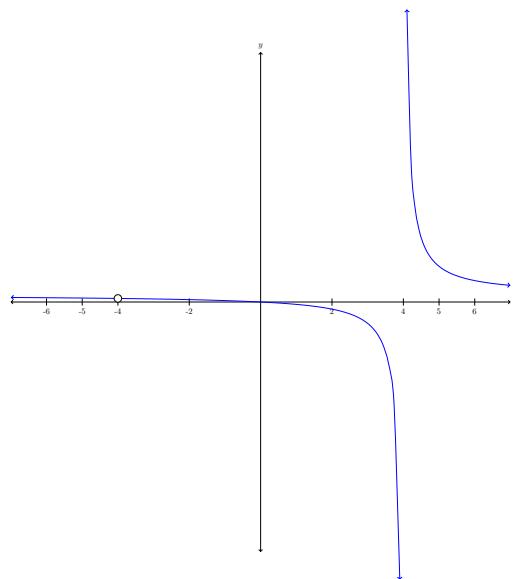


Figure 1.1.5: A plot of the function $y = f(x)$.

Solution. We first consider a graph of $y = f(x)$.

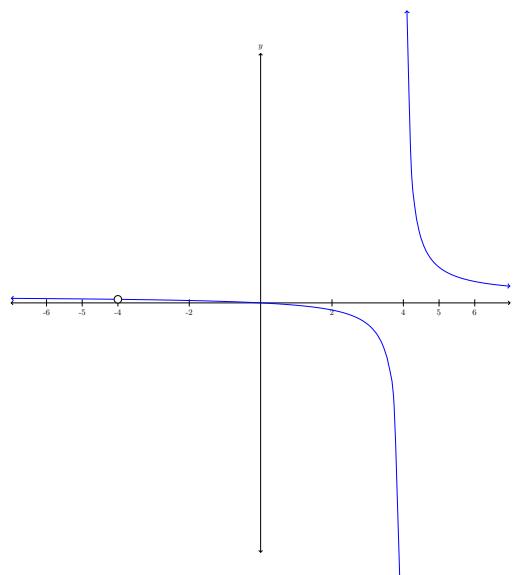


Figure 1.1.6: A plot of the function $y = f(x)$.

Visual inspection of the graph certainly indicates discontinuities. However, we can make our visual inspection more precise through a little algebra. We start by expanding the numerator

and denominator of the function. Specifically,

$$f(x) = \frac{3x(x+4)}{x^2 - 16} = \frac{3x(x+4)}{(x+4)(x-4)}.$$

From here we see that there are two x -values at which the function is undefined, $x = -4, 4$. Since the term $x+4$ occurs in the numerator and denominator $x = -4$ is a removable discontinuity. Since the term $x-4$ occurs only in the denominator, $x = 4$ is an infinite discontinuity.

Note 1.1.7. Rational functions can be a challenging topic. For a more complete discussion of rational functions we refer the reader to Short-Run Behavior of Rational Functions

1.1.2 Piecewise Functions

In many cases a simple function like $f(x) = x^2$ may not fully describe the behavior of a phenomena. In some of these cases we can turn to *piecewise* functions to give us the tools we need.

Piecewise Function

A piecewise-defined function is one which is defined not by a single equation, but by two or more. Each equation is valid for some interval .

Example 1.1.8. Graph the function defined by

$$f(x) = \begin{cases} x+1 & \text{if } x \leq 1 \\ 3 & \text{if } x > 1 \end{cases}.$$

Solution. Think of the plane as divided into two regions by the vertical line $x = 1$, as shown in Figure 1.1.9. In the left-hand region ($x \leq 1$), we graph the line $y = x + 1$. (The fastest way to graph the line is to plot its intercepts, $(-1, 0)$ and $(0, 1)$.)

Notice that the value $x = 1$ is included in the first region, so $f(1) = 1 + 1 = 2$, and the point $(1, 2)$ is included on the graph. We indicate this with a solid dot at the point $(1, 2)$.

In the right-hand region ($x > 1$), we graph the horizontal line $y = 3$. The value $x = 1$ is not included in the second region, so the point $(1, 3)$ is not part of the graph. We indicate this with an open circle at the point $(1, 3)$.

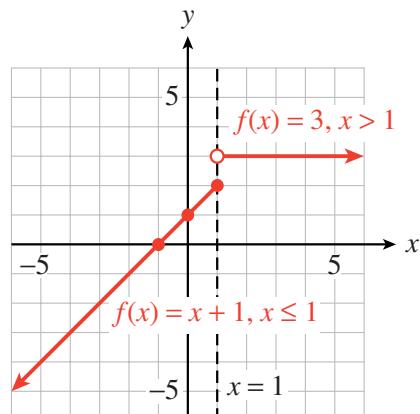


Figure 1.1.9

The jumps that a piecewise function possesses make piecewise functions a natural place in which to explore continuity.

Example 1.1.10. The function

$$h(x) = \begin{cases} 4x^2 & \text{if } x \leq 1 \\ 3x + 3 & \text{if } x > 1 \end{cases}$$

is discontinuous at $x = 1$. It is continuous on any interval that does not include $x = 1$.

Not only can we ask questions about when a piecewise function is continuous but we can also ask questions about how to make a piecewise function continuous by varying parameters.

Example 1.1.11. Consider the piecewise function

$$D(x) = \begin{cases} 4x^2 - k & \text{if } x < 2, \\ kx + 1 & \text{if } 2 \leq x. \end{cases}$$

Find the value of k to make this function continuous for all x .

Hint. The point where the function may not be continuous is at the potential "jump"-point. Specifically, when $x = 2$

Solution. The point where the function may not be continuous is at the potential "jump"-point. Specifically, when $x = 2$. For the function to be continuous we need the two "branches" of the function to meet when $x = 2$. That is we need $4 \times 2^2 - k = k \times 2 + 1$ which is equivalent to $k = 5$. The figure below shows a graph of $y = D(x)$ for $k = 5$.

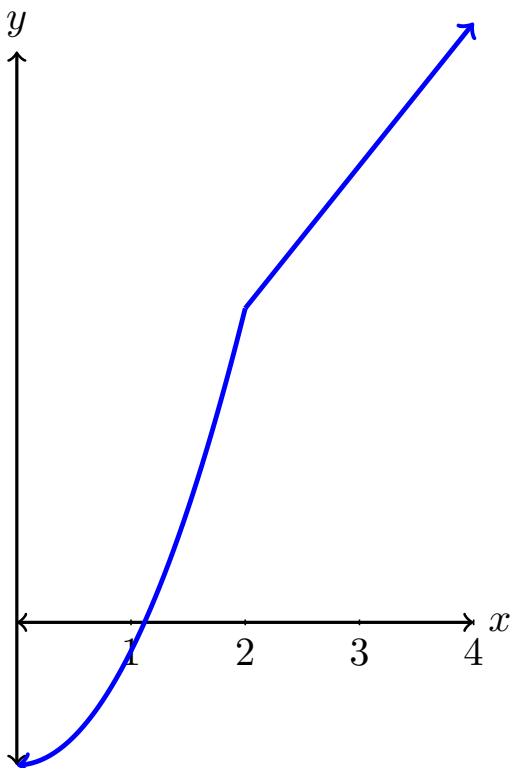


Figure 1.1.12: A plot of the function $y = f(x)$.

1.1.3 The Intermediate Value Theorem

Up to this point we have talked about the continuity of a function at a point. Before stating the intermediate value theorem we begin by defining a continuous function.

Continuous Function

A function is said to be continuous on an interval (a, b) if the function has no points of discontinuity on that interval.

One of the most important theorems in Calculus is the intermediate value theorem. The intermediate value theorem essentially says that a continuous function takes the value of any any points that it crosses. The formal statement of the theorem follows.

Intermediate Value theorem

Suppose f is continuous on a closed interval $[a, b]$. If k is any number between $f(a)$ and $f(b)$, then there is at least one number c in $[a, b]$ such that $f(c) = k$.

1.1.4 Exercises

1. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
9. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
10. This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

1.2 Introduction to Limits

Motivating Questions

- 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 $\lim_{x \rightarrow a} f(x) = L$?
 - How do we go about determining the value of the limit of a function at a point?
 - What is a left-hand limit at $x = a$ and a right-hand limit at $x = a$?
 - What does it mean graphically to say that f has limit L as $x \rightarrow a$? How is this connected to having a left-hand limit at $x = a$ and having a right-hand limit at $x = a$?
 - How are limits connected to continuity?
-

Limits are a mathematical construct we can use to describe the behavior of a function near a point. Consider the function g given by the graph in Figure 1.2.1 below. We can evaluate the function at a variety of points. For example, $g(-2) = 0$, $g(-1) = 3$, and $g(0) = 1$.

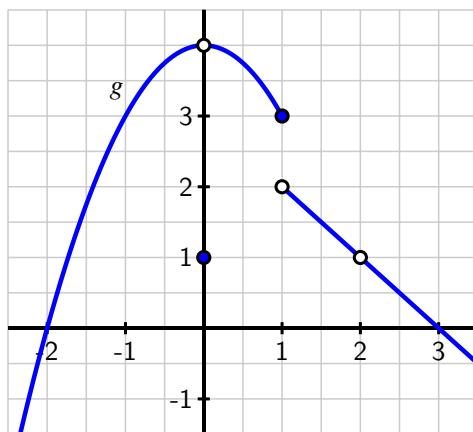


Figure 1.2.1: Graph of $y = g(x)$.

As stated earlier, $g(0) = 1$. However, a careful look at the graph shows that you might have expected $g(0)$ to be 4. In fact, you would probably agree that “As x gets closer and closer (but not equal) to 0, $g(x)$ gets as close as we want to 4.” This is the basic idea of a limit and we will expand on this idea in the following chapter.

1.2.1 The Notion of Limit

Limits give us a way to identify a trend in the values of a function as its input variable approaches a particular value of interest. We need a precise understanding of what it means to say “a function f has limit L as x approaches a .”

In Figure 1.2.1, we saw that 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. But limits describe the behavior of a function *arbitrarily close to* a fixed input and are not affected by the value of the function *at* the fixed input. More formally,¹ we say the following.

If a function f is defined on an interval around c , except perhaps at the point $x = c$. We define the limit of the function $f(x)$ as x approaches c to be a number L (if one exists) such that $f(x)$ is as close to L as we want whenever x is sufficiently close to c (but $x \neq c$). If L exists, we write

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

On the other hand, if as x approaches c we cannot make $f(x)$ as close to a single value as we would like, then we say that f **does not have a limit as x approaches c** .

Example 1.2.2. Recall the function g from the section introduction.

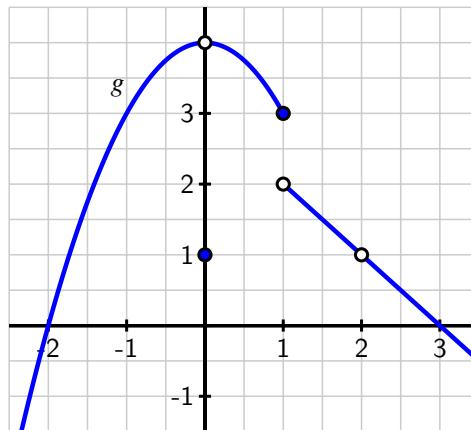


Figure 1.2.3: Graph of $y = g(x)$

For the function g pictured in Figure 1.2.3, we make the following observations:

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

¹What 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. That being said, the definition presented here is sufficient for the purposes of this text.

When finding a limit from a graph, it suffices to ask if the function *approaches* a single value from each side of the fixed input. The function value *at* the fixed input is irrelevant. This reasoning explains the values of the three limits stated above.

We further observe that g does not have a limit as x approaches 1 because there is a jump in the graph at $x = 1$. If we approach $x = 1$ from the left, the function values tend to get close to 3, but if we approach $x = 1$ from the right, the function values get close to 2. 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. 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 values as the input values approach a .

The first approach produces only an approximation of the value of the limit, while the latter can often be used to determine the limit exactly.

Example 1.2.4 (Limits of Two Functions). For each of the following functions, we'd like to know whether or not the function has a limit at the stated c -values. Use both numerical and algebraic approaches to investigate and, if possible, estimate or determine the value of the limit. Compare the results with a careful graph of the function on an interval containing the points of interest.

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

b. $g(x) = \sin\left(\frac{\pi}{x}\right); c = 3, c = 0$

Solution. a. We first construct a graph of f along with tables of values near $c = -1$ and $c = -2$.

x	$f(x)$	x	$f(x)$
-0.9	2.9	-1.9	3.9
-0.99	2.99	-1.99	3.99
-0.999	2.999	-1.999	3.999
-0.9999	2.9999	-1.9999	3.9999
-1.1	3.1	-2.1	4.1
-1.01	3.01	-2.01	4.01
-1.001	3.001	-2.001	4.001
-1.0001	3.0001	-2.0001	4.0001

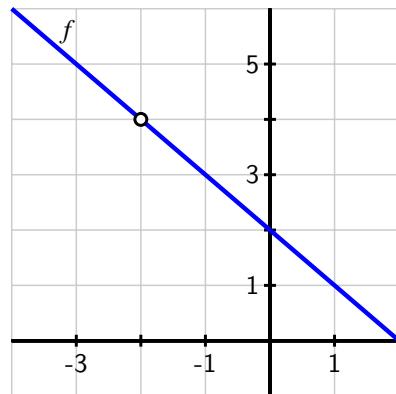


Table 1.2.5: Table of f values near $x = -1$.

Table 1.2.6: Table of f values near $x = -2$.

Figure 1.2.7: Plot of $f(x)$ on $[-4, 2]$.

From Table 1.2.5, it appears that we can make f as close as we want to 3 by taking x sufficiently close to -1 , which suggests that $\lim_{x \rightarrow -1} f(x) = 3$. This is also consistent with the graph of f , as seen in Figure 1.2.7. To see this a bit more rigorously and from an algebraic point of view, consider the formula for f : $f(x) = \frac{4-x^2}{x+2}$. As x approaches -1 , the numerator, $4-x^2$, of f approaches $4-(-1)^2 = 3$, and the denominator, $x+2$, of f approaches $-1+2 = 1$. Hence $\lim_{x \rightarrow -1} f(x) = \frac{3}{1} = 3$.

The situation is more complicated when x approaches -2 because $f(-2)$ is not defined. If we try to use a similar algebraic argument regarding the numerator and denominator, we observe that as x approaches -2 , the numerator $(4-x^2)$ approaches $(4-(-2)^2) = 0$, and the denominator $(x+2)$ approaches $(-2+2) = 0$, so as x approaches -2 , the numerator and denominator of f both tend to 0. We call $0/0$ an *indeterminate form*. This tells us that there is somehow more work to do. From Table 1.2.6 and Figure 1.2.7, it appears that f should have a limit of 4 at $x = -2$.

To see algebraically why this is the case, observe that

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

It is important to observe that because we are taking the limit as x approaches -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 find that

$$\lim_{x \rightarrow -2} f(x) = \lim_{x \rightarrow -2} (2-x).$$

This limit is now easy to determine, and its value is 4. Thus, from several points of view

we've seen that $\lim_{x \rightarrow -2} f(x) = 4$.

b. Next we turn to the function g , and construct two tables and a graph.

x	$g(x)$	x	$g(x)$
2.9	0.84864	-0.1	0
2.99	0.86428	-0.01	0
2.999	0.86585	-0.001	0
2.9999	0.86601	-0.0001	0
3.1	0.88351	0.1	0
3.01	0.86777	0.01	0
3.001	0.86620	0.001	0
3.0001	0.86604	0.0001	0

Table 1.2.8: Table of g values near $x = 3$.

Table 1.2.9: Table of g values near $x = 0$.

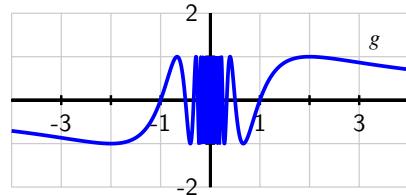


Figure 1.2.10: Plot of $g(x)$ on $[-4, 4]$.

First, as x approaches 3, it appears from the values in Table 1.2.8 that the function is approaching a number between 0.86601 and 0.86604. From the graph in Figure 1.2.10 it appears that $g(x)$ approaches $g(3) = \sin(\frac{\pi}{3})$ as x approaches 3. The exact value of $g(3) = \sin(\frac{\pi}{3})$ is $\frac{\sqrt{3}}{2}$, which is approximately 0.8660254038. This is convincing evidence that

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

As x approaches 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 increasingly rapid oscillations in the graph of g between 1 and -1. If we plot the function $g(x) = \sin(\frac{\pi}{x})$ with a graphing utility and then zoom in on $x = 0$, we see that the function never settles down to a single value near the origin, which suggests that g does not have a limit at $x = 0$.

How do we reconcile the graph in Figure 1.2.10 with Table 1.2.9, which seems to suggest that the limit of g as x approaches 0 may in fact be 0? The data misleads us because of the special nature of the sequence of input values $\{0.1, 0.01, 0.001, \dots\}$. When we evaluate $g(10^{-k})$, we get $g(10^{-k}) = \sin\left(\frac{\pi}{10^{-k}}\right) = \sin(10^k \pi) = 0$ for each positive integer value of k . But if we take a different sequence of values approaching zero, say $\{0.3, 0.03, 0.003, \dots\}$, then we find that

$$g(3 \cdot 10^{-k}) = \sin\left(\frac{\pi}{3 \cdot 10^{-k}}\right) = \sin\left(\frac{10^k \pi}{3}\right) = \frac{\sqrt{3}}{2} \approx 0.866025.$$

That sequence of function values suggests that the value of the limit is $\frac{\sqrt{3}}{2}$. Clearly the function cannot have two different values for the limit, so g has no limit as x approaches 0.

An important lesson to take from Example 1.2.4 is that 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.

Example 1.2.11. Estimate the value of each of the following limits by constructing appropriate tables of values. Then determine the exact value of the limit by using algebra to simplify the function. Finally, plot each function on an appropriate interval to check your result visually.

a. $\lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1}$

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

c. $\lim_{x \rightarrow 0} \frac{\sqrt{x+1} - 1}{x}$

Hint.

a. $(x^2 - 1)$ can be factored.

b. Expand the expression $(2 + x)^3$, and then combine like terms in the numerator.

c. Try multiplying the given function by this fancy form of 1: $\frac{\sqrt{x+1}+1}{\sqrt{x+1}+1}$.

Answer.

a. 2.

b. 12.

c. $\frac{1}{2}$.

Solution. Estimating the values of the limits with tables is straightforward and should suggest the exact values stated below.

a. $\lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1} = \lim_{x \rightarrow 1} \frac{(x+1)(x-1)}{x-1} = \lim_{x \rightarrow 1} (x+1) = 2.$

b.

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{(2+x)^3 - 8}{x} &= \lim_{x \rightarrow 0} \frac{8 + 12x + 6x^2 + x^3 - 8}{x} = \lim_{x \rightarrow 0} \frac{12x + 6x^2 + x^3}{x} \\ &= \lim_{x \rightarrow 0} (12 + 6x + x^2) = 12. \end{aligned}$$

c.

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\sqrt{x+1} - 1}{x} &= \lim_{x \rightarrow 0} \frac{\sqrt{x+1} - 1}{x} \cdot \frac{\sqrt{x+1} + 1}{\sqrt{x+1} + 1} = \lim_{x \rightarrow 0} \frac{x+1-1}{x(\sqrt{x+1} + 1)} \\ &= \lim_{x \rightarrow 0} \frac{1}{\sqrt{x+1} + 1} = \frac{1}{2}. \end{aligned}$$

Recall 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.

1.2.2 Having a limit at a point

We saw earlier that f has limit L as x approaches c provided that we can make the value of $f(x)$ as close to L as we like by taking x sufficiently close (but not equal to) c . If so, we write $\lim_{x \rightarrow c} f(x) = L$. We also saw that there are cases where a function can fail to have a limit. The graphs that follow are two such examples.

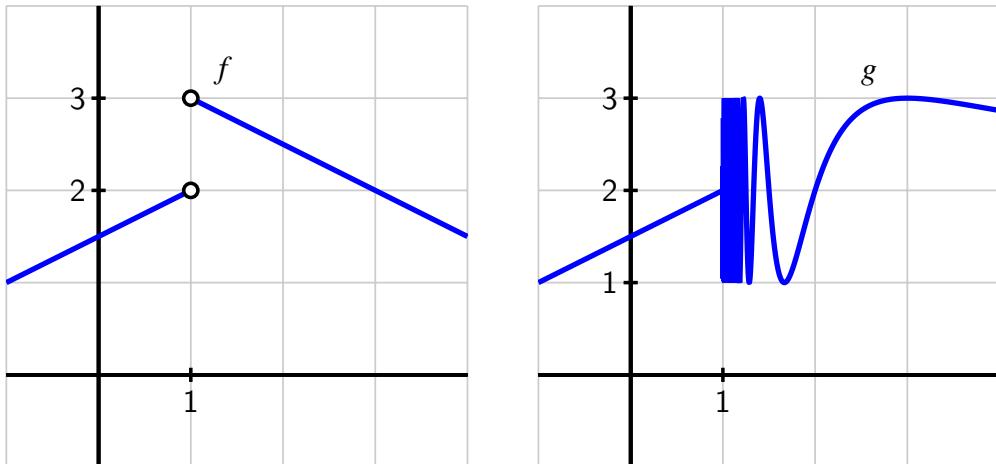


Figure 1.2.12: Functions f and g that each fail to have a limit at $c = 1$.

Essentially there are two behaviors that a function can exhibit near a point where it fails to have a limit. In Figure 1.2.12, at left we see a function f whose graph shows a jump at $c = 1$. If we let x approach 1 from the left side, the value of f approaches 2, but if we let x approach 1 from the right, the value of f tends to 3. Because the value of f does not approach a single number as x gets arbitrarily close to 1 from both sides, we know that f does not have a limit at $c = 1$.

For such cases, we introduce the notion of *left* and *right* (or *one-sided*) limits.

We say that f has limit L_1 as x approaches c from the left and write

$$\lim_{x \rightarrow c^-} f(x) = L_1$$

provided that we can make the value of $f(x)$ as close to L_1 as we like by taking x sufficiently close to c while always having $x < c$. We call L_1 the left-hand limit of f as x approaches c . Similarly, we say L_2 is the right-hand limit of f as x approaches c and write

$$\lim_{x \rightarrow c^+} f(x) = L_2$$

provided that we can make the value of $f(x)$ as close to L_2 as we like by taking x sufficiently close to c while always having $x > c$.

In the graph of the function f in Figure 1.2.12, we see that

$$\lim_{x \rightarrow 1^-} f(x) = 2 \text{ and } \lim_{x \rightarrow 1^+} f(x) = 3.$$

Precisely because the left and right limits are not equal, the overall limit of f as $x \rightarrow 1$ fails to exist.

For the function g pictured at right in Figure 1.2.12, the function fails to have a limit at $c = 1$ for a different reason. While the function does not have a jump in its graph at $c = 1$, it is still not the case that g approaches a single value as x approaches 1. In particular, due to the infinitely oscillating behavior of g to the right of $c = 1$, we say that the right-hand limit of g as $x \rightarrow 1^+$ does not exist, and thus $\lim_{x \rightarrow 1} g(x)$ does not exist.

To summarize, if either a left- or right-hand limit fails to exist or if the left- and right-hand limits are not equal to each other, the overall limit does not exist.

A function f has limit L as $x \rightarrow c$ if and only if

$$\lim_{x \rightarrow c^-} f(x) = L = \lim_{x \rightarrow c^+} f(x).$$

That is, a function has a limit at $x = c$ if and only if both the left- and right-hand limits at $x = c$ exist and have the same value.

The function f given in Figure 1.2.13 fails to have a limit at only two values: at $c = -2$ (where the left- and right-hand limits are 2 and -1 , respectively) and at $x = 2$, where $\lim_{x \rightarrow 2^+} f(x)$ does not exist. Note well that even at values such as $c = -1$ and $c = 0$ where there are holes in the graph, the limit still exists.

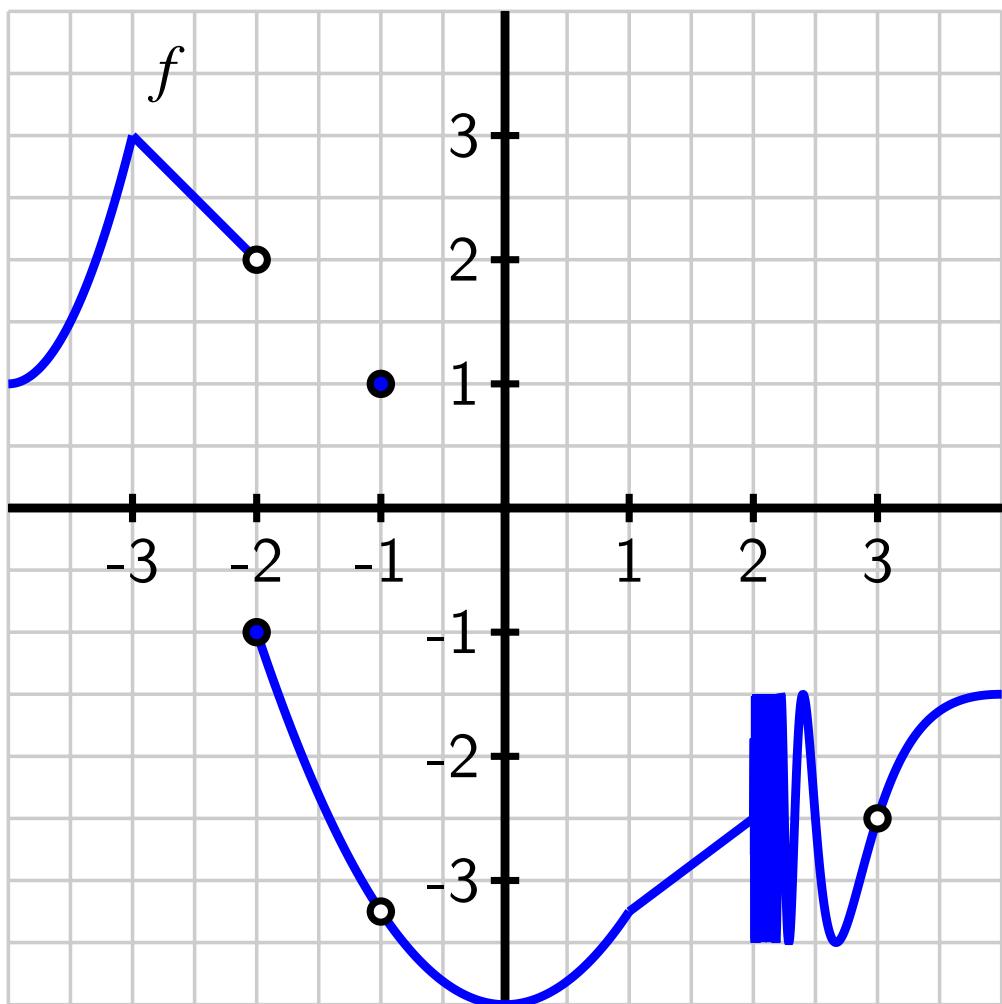


Figure 1.2.13: Functions f , g , and h that demonstrate subtly different behaviors at $c = 1$.

Example 1.2.14. Consider a function that is piecewise-defined according to the formula

$$f(x) = \begin{cases} 3(x + 2) + 2 & \text{for } -3 < x < -2 \\ \frac{2}{3}(x + 2) + 1 & \text{for } -2 \leq x < -1 \\ \frac{2}{3}(x + 2) + 1 & \text{for } -1 < x < 1 \\ 2 & \text{for } x = 1 \\ 4 - x & \text{for } x > 1 \end{cases}$$

Use the given formula to answer the following questions.

- For each of the values $c = -2, -1, 0, 1, 2$, compute $f(c)$.
- For each of the values $c = -2, -1, 0, 1, 2$, determine $\lim_{x \rightarrow c^-} f(x)$ and $\lim_{x \rightarrow c^+} f(x)$.
- For each of the values $c = -2, -1, 0, 1, 2$, determine $\lim_{x \rightarrow c} f(x)$. If the limit fails to exist, explain why by discussing the left- and right-hand limits at the relevant a -value.
- For which values of c is the following statement true?

$$\lim_{x \rightarrow c} f(x) \neq f(c)$$
- Sketch an accurate, labeled graph of $y = f(x)$. Be sure to carefully use open circles (\circ) and filled circles (\bullet) to represent key points on the graph, as dictated by the piecewise formula.

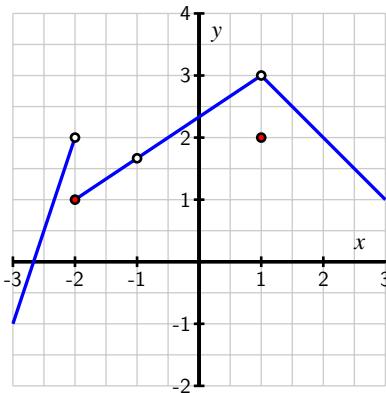
Hint.

- Find the interval in which c lies and evaluate the function there.
- Remember that for $\lim_{x \rightarrow c^-} f(x)$, we only consider values of x such that $x < c$. Find the right formula to use in the piecewise definition for f to fit the values you are considering.
- Use your work in (c) and compare left- and right-hand limits.
- Use your work in (a) and (c).
- Note that f is piecewise linear.

Answer.

- $f(-2) = 1$; $f(-1)$ is not defined; $f(0) = \frac{7}{3}$; $f(1) = 2$; $f(2) = 2$.
- $\lim_{x \rightarrow -2^-} f(x) = 2$ and $\lim_{x \rightarrow -2^+} f(x) = 1$
 $\lim_{x \rightarrow -1^-} f(x) = \frac{5}{3}$ and $\lim_{x \rightarrow -1^+} f(x) = \frac{5}{3}$
 $\lim_{x \rightarrow 0^-} f(x) = \frac{7}{3}$ and $\lim_{x \rightarrow 0^+} f(x) = \frac{7}{3}$
 $\lim_{x \rightarrow 1^-} f(x) = 3$ and $\lim_{x \rightarrow 1^+} f(x) = 3$
 $\lim_{x \rightarrow 2^-} f(x) = 2$ and $\lim_{x \rightarrow 2^+} f(x) = 2$
- $\lim_{x \rightarrow -2} f(x)$ does not exist. The values of the limits as $x \rightarrow c$ for $c = -1, 0, 1, 2$ are $\frac{5}{3}, \frac{7}{3}, 3, 2$.
- $c = -2, c = -1$, and $c = 1$.

e.

**Solution.**

- a. $f(-2) = \frac{2}{3}(-2 + 2) + 1 = 1$; $f(-1)$ is not defined; $f(0) = \frac{2}{3}(0 + 2) + 1 = \frac{7}{3}$; $f(1) = 2$ (by the rule); $f(2) = 4 - 2 = 2$.

b.

$$\lim_{x \rightarrow -2^-} f(x) = 2 \text{ and } \lim_{x \rightarrow -2^+} f(x) = 1$$

$$\lim_{x \rightarrow -1^-} f(x) = \frac{5}{3} \text{ and } \lim_{x \rightarrow -1^+} f(x) = \frac{5}{3}$$

$$\lim_{x \rightarrow 0^-} f(x) = \frac{7}{3} \text{ and } \lim_{x \rightarrow 0^+} f(x) = \frac{7}{3}$$

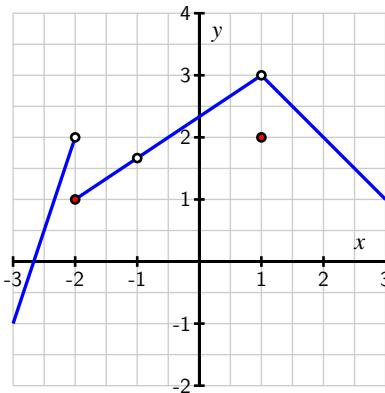
$$\lim_{x \rightarrow 1^-} f(x) = 3 \text{ and } \lim_{x \rightarrow 1^+} f(x) = 3$$

$$\lim_{x \rightarrow 2^-} f(x) = 2 \text{ and } \lim_{x \rightarrow 2^+} f(x) = 2$$

- c. $\lim_{x \rightarrow -2} f(x)$ does not exist because the left-hand limit is 2 while the right-hand limit is 1. All of the other requested limits exist, as left- and right-hand limits exist and are equal in each case. The respective values of the limits as $x \rightarrow c$ for $c = -1, 0, 1, 2$ are $\frac{5}{3}, \frac{7}{3}, 3, 2$.

- d. For $c = -2$, $a = -1$, and $c = 1$, $\lim_{x \rightarrow c} f(x) \neq f(a)$. At $c = -2$, the limit fails to exist, but $f(-2) = 1$. At $c = -1$, the limit is $\frac{5}{3}$, but $f(-1)$ is not defined. At $c = 1$, the limit is 3, but $f(1) = 2$.

e.



1.2.3 Limits and Continuity

Recall that informally we said a function is continuous if we can draw its graph without ever lifting our pencil from the page.

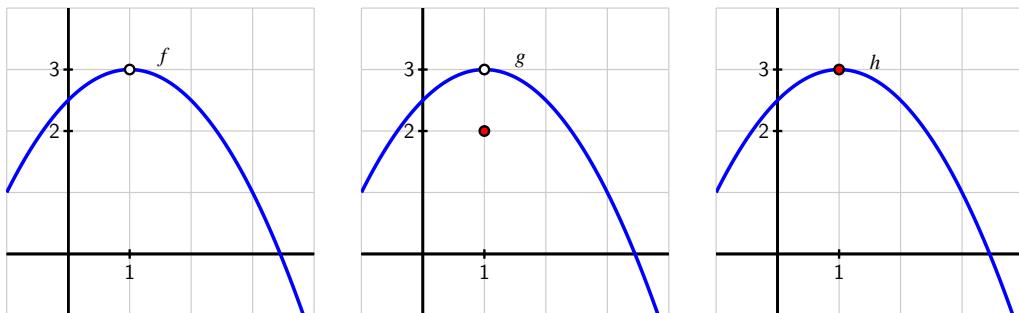


Figure 1.2.15: Functions f , g , and h that demonstrate subtly different behaviors at $c = 1$.

Using limits we can formalize the idea of continuity. First consider the function in the left-most graph. Note that $f(1)$ is not defined, which leads to the resulting hole in the graph of f at $c = 1$. We will naturally say that f is *not continuous* at $c = 1$. For the function g , we observe that while $\lim_{x \rightarrow 1} g(x) = 3$, 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 $c = 1$. Finally, the function h appears to be the most well-behaved of all three, since at $c = 1$ its limit and its function value agree. That is,

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

With no hole or jump in the graph of h at $c = 1$, we say that h is *continuous* there. More formally, we make the following definition.

A function f is **continuous at $x = c$** provided that

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

Conditions (a) and (b) are technically contained implicitly in (c), but we state them explicitly to emphasize their individual importance. The definition says that a function is continuous at $x = c$ provided that its limit as $x \rightarrow c$ exists and equals its function value at $x = c$. If a function is continuous at every point in an interval $[a, b]$, we say the function is “continuous on $[a, b]$.” If a function is continuous at every point in its domain, we simply say the function is “continuous.” 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.

For example, 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 $\lim_{x \rightarrow 2} p(x)$, we simply compute

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

This route of substituting an input value to evaluate a limit works whenever we know that the 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 1.2.16. Using the same function f as given by the graph that is in Figure 1.2.17.

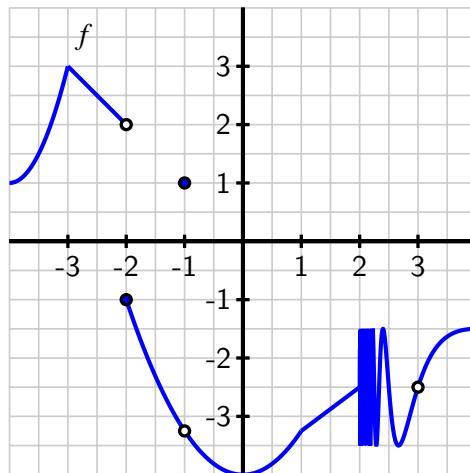


Figure 1.2.17: The graph of $y = f(x)$ for Example 1.2.16.

- At which values of c does $\lim_{x \rightarrow c} f(x)$ not exist?
- At which values of c is $f(c)$ not defined?
- At which values of c does f have a limit, but $\lim_{x \rightarrow c} f(x) \neq f(c)$?
- State all values of c for which f is not continuous at $x = c$.
- Which condition is stronger, and hence implies the other: f has a limit at $x = c$ or f is continuous at $x = c$? Explain, and hence complete the following sentence: "If f _____ at $x = c$, then f _____ at $x = c$," where you complete the blanks with *has a limit* and *is continuous*, using each phrase once.

Hint.

- Consider the left- and right-hand limits at each value.
- Carefully examine places on the graph where there's an open circle.
- Are there locations on the graph where the function has a limit but there's a hole in the graph?
- Remember that at least one of three conditions must fail: if the function lacks a limit, if the function is undefined, or if the limit exists but does not equal the function value, then f is not continuous at the point.
- Note that the definition of being continuous requires the limit to exist.

Answer.

- $c = -2; c = +2$.
- $c = 3$.
- $c = -1; c = 3$.
- $c = -2; c = 2; c = 3; c = -1$.
- "If f is continuous at $x = c$, then f has a limit at $x = c$."

Solution.

- $\lim_{x \rightarrow c} f(x)$ does not exist at $c = -2$ since $\lim_{x \rightarrow -2^-} f(x) = 2 \neq -1 = \lim_{x \rightarrow -2^+}$ and $\lim_{x \rightarrow c} f(x)$ does not exist at $c = +2$ since $\lim_{x \rightarrow 2^+} f(x)$ does not exist due to the infinitely oscillatory behavior of f .
- The only point at which f is not defined is at $c = 3$.
- At $c = -1$, note that $\lim_{x \rightarrow -1} f(x)$ exists (and appears to have value approximately -3.25), but $f(-1) = 1$, and thus $\lim_{x \rightarrow -1} f(x) \neq f(-1)$. At $c = 3$, $\lim_{x \rightarrow 3} f(x) = -2.5$, but $f(3)$ is not defined, so the limit exists but does not equal the function value.

- d. Based on our work in (a), (b), and (c), f is not continuous at $c = -2$ and $c = 2$ because f does not have a limit at those points; f is not continuous at $c = 3$ since f is not defined there; and f is not continuous at $c = -1$ because at that point its limit does not equal its function value.
- e. “If f is continuous at $x = c$, then f has a limit at $x = c$,” since one of the defining properties of “being continuous” at $x = c$ is that the function has a limit at that input value. This shows that being continuous is a stronger condition than having a limit.

1.2.4 Properties of Limits and Continuous Functions

There are several properties of limits and continuous functions that are useful to have in your toolbox. Specifically, limits and continuous functions behave well under typical mathematical operations. While these properties can be proven we proceed to only state the properties.

Properties of Limits

Assuming all the limits on the right-hand side exist:

- If b is a constant, then $\lim_{x \rightarrow c} (bf(x)) = b \left(\lim_{x \rightarrow c} f(x) \right)$
- $\lim_{x \rightarrow c} (f(x) + g(x)) = \lim_{x \rightarrow c} f(x) + \lim_{x \rightarrow c} g(x)$
- $\lim_{x \rightarrow c} (f(x) \cdot g(x)) = \lim_{x \rightarrow c} f(x) \cdot \lim_{x \rightarrow c} g(x)$
- $\lim_{x \rightarrow c} \left(\frac{f(x)}{g(x)} \right) = \frac{\lim_{x \rightarrow c} f(x)}{\lim_{x \rightarrow c} g(x)}$, provided $\lim_{x \rightarrow c} g(x) \neq 0$
- For any constant k , $\lim_{x \rightarrow c} k = k$
- $\lim_{x \rightarrow c} x = c$

Example 1.2.18. We can use algebra to compute $\lim_{x \rightarrow 1} x^2(x^3 + 2)$. Specifically,

$$\begin{aligned}
 \lim_{x \rightarrow 1} x^2(x^3 + 2) &= \left(\lim_{x \rightarrow 1} x^2 \right) \left(\lim_{x \rightarrow 1} (x^3 + 2) \right) \\
 &= \left(\lim_{x \rightarrow 1} x^2 \right) \left(\lim_{x \rightarrow 1} x^3 + \lim_{x \rightarrow 1} 2 \right) \\
 &= 1(1 + 2) \\
 &= 3.
 \end{aligned}$$

Continuity of Sums, Products, and Quotients of Functions

Suppose that f and g are continuous on an interval and that b is a constant. Then, on that same interval,

- $bf(x)$ is continuous.

- $f(x) + g(x)$ is continuous.
- $f(x)g(x)$ is continuous.
- $f(x)/g(x)$ is continuous, provided $g(x) \neq 0$ on the interval.

Continuity of Composite Functions

If f and g are continuous, and if the composite function $f(g(x))$ is defined on an interval, then $f(g(x))$ is continuous on that interval.

1.2.5 Exercises

1. (Limits on a piecewise graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Estimating a limit numerically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Limits for a piecewise formula) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Limits for a piecewise formula) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Calculating Limits, Rational Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (One Sided Limits) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Evaluating a limit algebraically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. (Limit Rules) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

1.3 How do we measure velocity?

Motivating Questions

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

Calculus can be viewed broadly as the study of change. A natural and important question to ask about any changing quantity is “how fast is the quantity changing?”

We begin with a simple problem: a ball is tossed straight up in the air. How is the ball moving? Questions like this one are central to our study of differential calculus.

Example 1.3.1. Suppose that the height s of a ball at time t (in seconds) is given by the formula $s(t) = 64 - 16(t - 1)^2$ (in feet).

- Construct a graph of $y = s(t)$ on the time interval $0 \leq t \leq 3$. Label at least six distinct points on the graph, including the three points showing when the ball was released, when the ball reaches its highest point, and when the ball lands.
- Describe the behavior of the ball on the time interval $0 < t < 1$ and on the time interval $1 < t < 3$. What occurs at the instant $t = 1$?
- 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 on the graph? What does this value tell us about the motion of the ball? In particular, what are the units on $AV_{[0.5,1]}$?

Hint.

- s is a quadratic function, so its graph is a parabola. Where is the vertex? What are the t -intercepts?
- Where is the vertex?
- How does this value relate to the points $(1, s(1))$ and $(0.5, s(0.5))$ on the graph? What are the units of $s(t)$? What are the units of t ? How might these units inform the units of $AV_{[0.5,1]}$?

Answer.

a.

Geogebra: www.geogebra.org/m/cqs42zua

Figure 1.3.2: A graph of $y = s(t)$ on the time interval $0 \leq t \leq 3$. You can drag the point to check your points.

- b. The ball reaches its highest point at $t = 1$, rising until then and falling after.
- c. $AV_{[0.5,1]} = 8$ ft/s measures the slope of the line between $(0.5, s(0.5))$ and $(1, s(1))$, and is the average velocity of the ball between $t = 0.5$ seconds and $t = 1$ second.

Solution.

a.

Geogebra: www.geogebra.org/m/cqs42zua

Figure 1.3.3: A graph of $y = s(t)$ on the time interval $0 \leq t \leq 3$. You can drag the point to check your points.

- b. The ball is rising on the interval $0 < t < 1$ and falling on the interval $1 < t < 3$. At the instant $t = 1$, the ball reaches its highest point, stops, and changes direction.
- c. We first compute $s(1) = 64 - 16(1 - 1)^2 = 64$ and $s(0.5) = 64 - 16(0.5 - 1)^2 = 60$. With these values, we can say

$$\begin{aligned} AV_{[0.5,1]} &= \frac{s(1) - s(0.5)}{1 - 0.5} \\ &= \frac{64 - 60}{0.5} \\ &= \frac{4}{0.5} \\ &= 8 \end{aligned}$$

Notice that in the numerator, we have the values $s(1)$ and $s(0.5)$, both measured in feet. In the denominator, we have the corresponding time values $t = 1$ and $t = 0.5$, both measured in seconds. The units of the quotient, then, should be feet per second. These are units of velocity, and in fact the value of $AV_{[0.5,1]}$ is the average velocity on the time interval $0.5 \leq t \leq 1$. Moreover, the formula for $AV_{[0.5,1]}$ is precisely the slope formula between the points $(0.5, s(0.5))$ and $(1, s(1))$.

1.3.1 Position and average velocity

The *position* of any moving object can be considered as a function of *time*. When the motion is along a straight line, the position is given by a single variable, which we denote by $s(t)$. For example, $s(t)$ might give the mile marker of a car traveling on a straight highway at time t in

hours. Similarly, the function $s(t)$ described in Example 1.3.1 is a position function, where the position of a ball is measured vertically relative to the ground.

On any time interval, a moving object also has an *average velocity*. For example, to compute a car's average velocity we divide the number of miles traveled by the time elapsed, which gives the velocity in *miles per hour*. Similarly, the value of $AV_{[0.5,1]}$ in Example 1.3.1 gave the average velocity of the ball on the time interval $[0.5, 1]$, measured in feet per second.

In general, following the notation we introduced earlier, we make the following definition:

Average Velocity

For an object moving in a straight line with position 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: the units on $AV_{[a,b]}$ are “units of s per unit of t ,” such as “miles per hour” or “feet per second.”

Example 1.3.4. The following questions concern the position function given by $s(t) = 64 - 16(t - 1)^2$, considered in Example 1.3.1.

- 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 graph provided below in Figure 1.3.5, 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 think the velocity of the ball at the instant $t = 0.8$ is? Why?

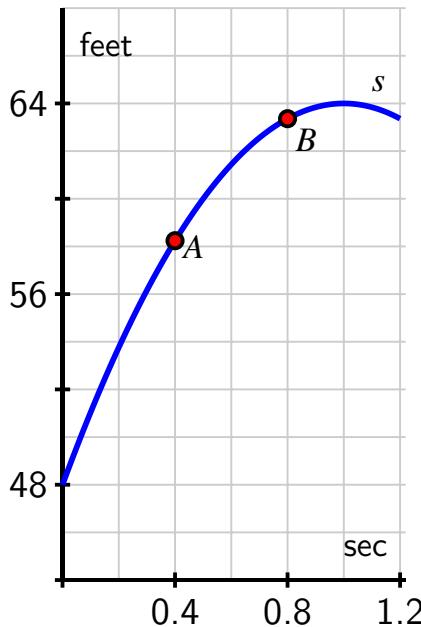


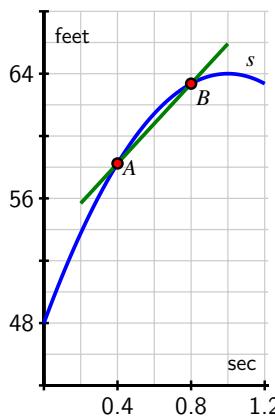
Figure 1.3.5: A partial plot of $s(t) = 64 - 16(t - 1)^2$.

Hint.

- On $[0.4, 0.8]$, the average velocity is $AV_{[0.4, 0.8]} = \frac{s(0.8) - s(0.4)}{0.8 - 0.4}$ ft/sec.
- Remember that the slope of a line can be found by taking “rise over run.” In this context, the slope is found by computing “change in s over change in t .”
- While the curve $s(t)$ is a parabola, how does it look up close on a very small interval?
- “Instantaneous” velocity can be approximated by average velocity on a very small interval.

Answer.

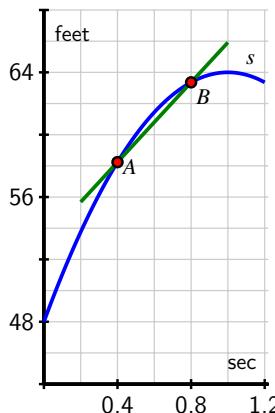
- $AV_{[0.4, 0.8]} = 12.8$ ft/sec; $AV_{[0.7, 0.8]} = 8$ ft/sec; the other average velocities are, respectively, 6.56, 6.416, 0, 4.8, 6.24, 6.384, all in ft/sec.
- $m = 12.8$ is the average velocity of the ball between $t = 0.4$ and $t = 0.8$.



- c. Like a straight line with slope about 6.4.
- d. About 6.4 feet per second.

Solution.

- a. On $[0.4, 0.8]$, the average velocity is $AV_{[0.4, 0.8]} = \frac{s(0.8) - s(0.4)}{0.8 - 0.4} = \frac{63.36 - 58.24}{0.4} = 12.8$ ft/sec. On $[0.7, 0.8]$, the average velocity is 8 ft/sec. The other average velocities are, respectively (in the order of the intervals listed in the activity), 6.56, 6.416, 0, 4.8, 6.24, 6.384, all measured in feet per second.
- b. The slope of the line between $A(0.4, s(0.4))$ and $B(0.8, s(0.8))$ is $\frac{s(0.8) - s(0.4)}{0.8 - 0.4} = 12.8$. This is precisely the average velocity of the ball between $t = 0.4$ and $t = 0.8$, and indeed each of the average velocities computed in (a) can be viewed as the slope of the line joining the points $(a, s(a))$ and $(b, s(b))$.



- c. As we zoom in on the curve $s(t) = 64 - 16(t - 1)^2$ at the point $(0.5, 60)$, the graph begins to look like a straight line. Indeed, it appears to look like a straight line with slope about 6.4.

- d. Observe that the average velocity of the ball on the intervals $[0.799, 0.8]$ and $[0.8, 0.801]$ is 6.416 and 6.384 feet/sec respectively. Hence it appears that the ball's velocity at the instant $t = 0.8$ should be about 6.4 feet per second.

1.3.2 Instantaneous Velocity

Whether we are driving a car, riding a bike, or throwing a ball, we have an intuitive sense that a 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 the car's velocity at any given instant. In fact, the velocity on a speedometer is really an average velocity that is computed over a very small time interval. If we let the time interval over which average velocity is computed become shorter and shorter, 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 $t = a$. We will develop a more formal definition of instantaneous velocity soon, and this definition will be the foundation of much of our work in calculus. For now, it is fine to think of instantaneous velocity as follows: take average velocities on smaller and smaller time intervals around a specific point. If those average velocities approach a single number, then that number will be the instantaneous velocity at that point.

Example 1.3.6. Each of the following questions concern $s(t) = 64 - 16(t - 1)^2$, the position function from Example 1.3.1.

- 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 $[0, 0.5]$?
- 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?
- How is the sign of the instantaneous velocity of the ball related to its behavior at a given point in time? In other words, what does positive instantaneous velocity tell you the ball is doing? Negative instantaneous velocity?
- Without doing any computations, what do you expect the instantaneous velocity of the ball at $t = 1$ to be? Why?

Hint.

- Remember to use the formula for average velocity from above: $AV_{[a,b]} = \frac{s(b)-s(a)}{b-a}$. Think carefully about whether certain quantities are positive or negative.
- To estimate the instantaneous velocity at $t = 1.5$, consider average velocities on the intervals $[1.499, 1.5]$ and $[1.5, 1.501]$.
- Think about whether the ball is rising or falling.

- d. What is the average velocity of the ball on small intervals that contain $t = 1$?

Answer.

- a. $AV_{[1.5,2]} = -24$ ft/sec, which is negative; $AV_{[0,0.5]} = 24$ ft/sec, which is positive.
- b. The instantaneous velocity at $t = 1.5$ is approximately -16 ft/sec; at $t = 2$, the instantaneous velocity is about -32 ft/sec, and $-16 > -32$.
- c. When the ball is rising its instantaneous velocity is positive, while when the ball is falling its instantaneous velocity is negative.
- d. Zero.

Solution.

- a. $AV_{[1.5,2]} = \frac{s(2)-s(1.5)}{2-1.5} = -24$ ft/sec. We note that this average velocity is negative, and in fact is the opposite of the average velocity of 24 ft/sec on the interval $[0, 0.5]$.
- b. Since $AV_{[1.499,1.5]} = -15.984$ and $AV_{[1.5,1.501]} = -16.016$, it appears that the instantaneous velocity of the ball at $t = 1.5$ is approximately -16 ft/sec. Similar computations show that at $t = 2$, the instantaneous velocity is about -32 ft/sec. Note that $-16 > -32$, so the instantaneous velocity at $t = 1.5$ is greater because it is “less negative.” Asking which number is “greater” is different from asking which number is “more negative.”
- c. When the ball is rising its instantaneous velocity is positive, while when the ball is falling its instantaneous velocity is negative.
- d. Note that $(1, s(1))$ is the vertex of the parabola given by $s(t)$. At this point, the ball is neither rising nor falling. On intervals of the form $[a, 1]$, where $a < 1$, the average velocity of the ball is positive; on intervals of form $[1, b]$, where $b > 1$, the average velocity is positive. Hence we expect the instantaneous velocity of the ball at the moment $t = 1$ to be zero.

At this point we have started to see a close connection between average velocity and instantaneous velocity. 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. We are interested in computing average velocities on the interval $[a, b]$ for smaller and smaller intervals. In order to make the link between average and instantaneous velocity more formal, think of the value b as $b = a + h$, where h is a small (non-zero) number that is allowed to vary. Then 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 $(a+h)-a = h$. Note that when $h < 0$, $AV_{[a,a+h]}$ measures the average velocity on the interval $[a+h, a]$.

To find the instantaneous velocity at $t = a$, we investigate what happens as the value of h approaches zero.

Example 1.3.7 (Computing instantaneous velocity for a falling ball). The position function for a falling ball is given by $s(t) = 16 - 16t^2$ (where $s(t)$ 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 $[0.4, 0.5]$.
- Make a conjecture about the instantaneous velocity at $t = 0.5$.

Hint.

- What formula do we have for the average velocity of a function f on an interval $[a, b]$? Can this be simplified at all with this function and interval? Why is this range of h values specified?
- What values of h do we need to use to get these intervals? Be careful with whether you choose h to be positive or negative.
- How might we use the expression we found for average velocity to estimate instantaneous velocity?

Answer.

- $AV_{[0.5, 0.5+h]} = -16 - 16h$.
- Use $h = 0.25$ to find $AV_{[0.5, 0.75]} = -20$ ft/sec, and $h = -0.1$ to find $AV_{[0.4, 0.5]} = -14.4$ ft/sec.
- The instantaneous velocity at $t = 0.5$ should be -16 ft/sec.

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. We want to compute and simplify

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

We start by finding $s(0.5 + h)$. To do so, we follow the rule that defines the function s .

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

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 remove the common factor of h from the numerator and denominator. It follows that

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

- b From this expression we can compute the average for any small positive or negative value of h . 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 $h = -0.1$, and compute the average velocity as

$$-16 - 16(-0.1) = -14.4 \text{ ft/sec.}$$

- c 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, so it appears that the instantaneous velocity of the ball at $t = 0.5$ should be -16 ft/sec.

Example 1.3.8. For the function given by $s(t) = 64 - 16(t - 1)^2$ from Example 1.3.1, 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 this to your earlier work in Example 1.3.4.

Hint. Note that

$$\begin{aligned} s(2+h) &= 64 - 16(2+h-1)^2 = 64 - 16(1+h)^2 \\ &= 64 - (16 + 32h + 16h^2) = 48 - 32h - 16h^2. \end{aligned}$$

Answer. $AV_{[2,2+h]} = -32 - 16h$

Solution. Observe first that

$$\begin{aligned} s(2+h) &= 64 - 16(2+h-1)^2 = 64 - 16(1+h)^2 \\ &= 64 - (16 + 32h + 16h^2) = 48 - 32h - 16h^2. \end{aligned}$$

Next, recall that $AV_{[2,2+h]} = \frac{s(2+h)-s(2)}{h}$, so

$$AV_{[2,2+h]} = \frac{s(2+h) - s(2)}{h} = \frac{(48 - 32h - 16h^2) - 48}{h} = \frac{-32h - 16h^2}{h}.$$

Now, since we assume $h \neq 0$, we can simplify further to find that $AV_{[2,2+h]} = -32 - 16h$. Setting $h = -0.5$, it follows that $AV_{[1.5,2]} = -32 + 16(0.5) = -24$ ft/sec. Likewise, as we let h approach zero, we see that $-32 - 16h$ will approach -32 , so the instantaneous velocity at $t = 2$ appears to be -32 feet/sec. Both results match our earlier work in Example 1.3.4.

1.3.3 Summary

- For an object moving in a straight line with position 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

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

- 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 in Figure 1.3.9.

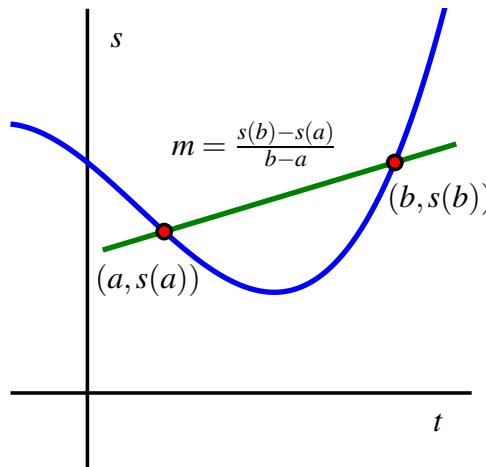


Figure 1.3.9: The graph of position function $s(t)$ together with the line through $(a, s(a))$ and $(b, s(b))$ whose slope is $m = \frac{s(b)-s(a)}{b-a}$. The line's slope is the average rate of change of $s(t)$ on the interval $[a, b]$.

- Given a moving object whose position at time t is given by a function $s(t)$, 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.

1.3.4 Exercises

1. (Average velocity from position) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Rate of calorie consumption) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Average rate of change - quadratic function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Comparing average rate of change of two functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Matching a distance graph to velocity) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. 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 1.3.10. 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 $s(t) = 100 \cos(0.75t) \cdot e^{-0.2t} + 100$.

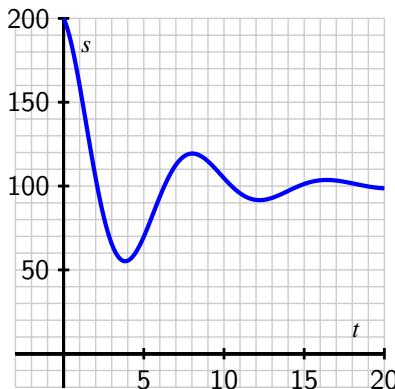


Figure 1.3.10: A bungee jumper's height function.

- What is the change in vertical position of the bungee jumper between $t = 0$ and $t = 15$?
- 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.
- On what time interval(s) do you think the bungee jumper achieves her greatest average velocity? Why?
- Estimate the jumper's instantaneous velocity at $t = 5$. Show your work and explain your reasoning, and include units on your answer.
- Among the average and instantaneous velocities you computed in earlier questions, which are positive and which are negative? What does negative velocity indicate?

Answer.

- a. $s(15) - s(0) \approx -98.75$.
 b.

$$AV_{[0,15]} = \frac{s(15) - s(0)}{15 - 0} \approx -6.58$$

$$AV_{[0,2]} = \frac{s(2) - s(0)}{2 - 0} \approx -47.63$$

$$AV_{[1,6]} = \frac{s(6) - s(1)}{6 - 1} \approx -13.25$$

$$AV_{[8,10]} = \frac{s(10) - s(8)}{10 - 8} \approx -7.35$$

- c. Most negative average velocity on $[0, 4]$; most positive average velocity on $[4, 8]$.
 d. $\frac{21.31+22.25}{2} = 21.78$ feet per second.
 e. The average velocities are negative; the instantaneous velocity was positive. Downward motion corresponds to negative average velocity; upward motion to positive average velocity.

Solution.

- a. Since $s(0) = 100 \cos(0)e^0 + 100 = 200$ and $s(15) = 100 \cos(11.25)e^{-3} + 100 \approx 101.25$, the bungee jumper's change in vertical position between $t = 0$ to $t = 15$ is

$$s(15) - s(0) \approx -98.75.$$

- b. We estimate the jumper's average velocity using the formula

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

Using the approximations from part (a) and $s(1) \approx 159.91$, $s(2) \approx 104.74$, $s(6) \approx 93.65$, $s(8) \approx 119.39$, $s(10) \approx 104.69$, we have

$$AV_{[0,15]} = \frac{s(15) - s(0)}{15 - 0} \approx -6.58$$

$$AV_{[0,2]} = \frac{s(2) - s(0)}{2 - 0} \approx -47.63$$

$$AV_{[1,6]} = \frac{s(6) - s(1)}{6 - 1} \approx -13.25$$

$$AV_{[8,10]} = \frac{s(10) - s(8)}{10 - 8} \approx -7.35$$

The units on each of these average velocities are "feet per second".

- c. On the time interval $[0, 4]$ the jumper has the greatest change in position over that period of time, so it is on that interval that her average velocity is greatest in magnitude (but negative in sign). For the same reasons, the largest (positive) average velocity will occur on the interval $[4, 8]$.
 d. We can approximate the jumper's average velocity on the intervals $[4.9, 5]$ and $[5, 5.1]$ and find the average of these average velocities. This will give us the slope of the secant line connecting the points $(4.9, s(4.9))$ and $(5.1, s(5.1))$ and be a reasonably good

approximation to the instantaneous velocity at $t = 5$. Now

$$AV_{[4.9,5]} = \frac{s(5) - s(4.9)}{5 - 4.9} \approx 21.31$$

$$AV_{[5,5.1]} = \frac{s(5.1) - s(5)}{5.1 - 5} \approx 22.25$$

so the jumper's instantaneous velocity at $t = 5$ is about

$$\frac{21.31 + 22.25}{2} = 21.78$$

feet per second.

- e. The average velocities we calculated were all negative, while the instantaneous velocity was positive. When the jumper is moving in the downward direction, the average velocity will be negative (indicating a decrease in position), while a positive average velocity indicates an increase in position (moving upwards).
- 7. 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 1.3.11. Include scale and units on the vertical axis. Be as detailed as possible.

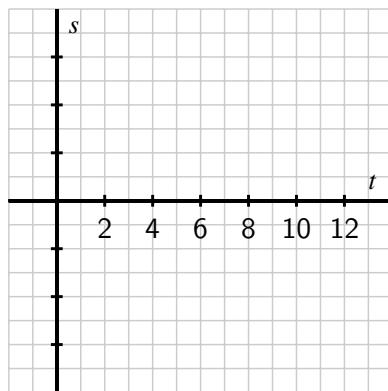


Figure 1.3.11: Axes for plotting $s(t)$ in part (a).

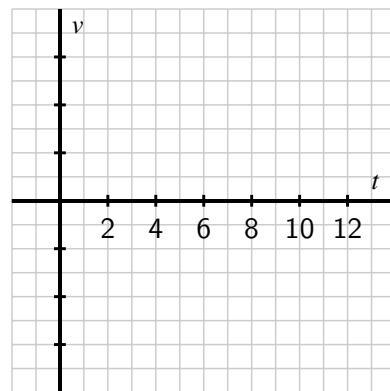


Figure 1.3.12: Axes for plotting $v(t)$ in part (c).

- 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 1.3.12. 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.

Answer.

- a. Sketch a plot where the diver's height at time t is on the vertical axis. For instance, $h(2.45) = 0$.
- b. $AV_{[2.45,7]} \approx \frac{-3.5-0}{7-2.45} = \frac{-3.5}{4.55} = -0.7692$ m/sec. The average velocity is not the same on every time interval within $[2.45, 7]$.
- c. When the diver is going upward, her velocity is positive. When she is going downward, her velocity is negative. At the peak of her dive and when her feet touch the bottom of the pool.
- d. It looks like when the position function is steep, the velocity function's value is farther away from zero, and that whenever the height/position function is rising/increasing, the velocity function has a positive value. Similarly, whenever the position function is decreasing, the velocity is negative.

Solution.

- a. Taking into account the given information and our intuition about the diver's velocity, sketch a plot where the diver's height at time t is on the vertical axis. For instance, $h(2.45) = 0$.
- b. First, we see that
- $$AV_{[2.45,7]} \approx \frac{-3.5-0}{7-2.45} = \frac{-3.5}{4.55} = -0.7692$$
- m/sec. The average velocity is not the same on every time interval within $[2.45, 7]$, since the position function is not linear. Physically, we expect the water to make the diver's velocity change as she glides through the water.
- c. When the diver is going upward, her velocity is positive. When she is going downward, her velocity is negative. At the peak of her dive and when her feet touch the bottom of the pool, her velocity is zero because the velocity is changing sign there. Sketch a possible graph accordingly.
- d. It looks like when the position function is steep, the velocity function's value is farther away from zero, and that whenever the height/position function is rising/increasing, the velocity function has a positive value. Similarly, whenever the position function is decreasing, the velocity is negative.
8. 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 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?

Answer.

- a. 15957 people.
- b. In an average year the population grew by about 798 people/year.
- c. The slope of a secant line through the points $(a, f(a))$ and $(b, f(b))$.
- d. $AV_{[0,20]} \approx 798$ people per year.
- e.

$$AV_{[5,10]} \approx 734.50$$

$$AV_{[5,9]} \approx 733.06$$

$$AV_{[5,8]} \approx 731.62$$

$$AV_{[5,7]} \approx 730.19$$

$$AV_{[5,6]} \approx 728.7535$$

Solution.

- a. The change in population is $P(2000) - P(1980) = 197800 - 181843 = 15957$ people.
- b. Since there were 15957 people added to Grand Rapids over the 20 years from 1980 to 2000, in an average year the population grew by $\frac{15,957}{20} \approx 798$ people/year.

- c. $\frac{f(b)-f(a)}{b-a} = \frac{\Delta f}{\Delta x}$ measures the slope of a secant line through the points $(a, f(a))$ and $(b, f(b))$.

d.

$$AV_{[0,20]} = \frac{P(20) - P(0)}{20 - 0} = \frac{197800 - 181843}{20} = \frac{15957}{20} \approx 798$$

people per year.

e.

$$AV_{[5,10]} = \frac{P(10) - P(5)}{10 - 5} = \frac{189116.72 - 185444.20}{10 - 5} \approx 734.50$$

$$AV_{[5,9]} = \frac{188376.44 - 185444.20}{9 - 5} \approx 733.06$$

$$AV_{[5,8]} = \frac{187639.06 - 185,444.20}{8 - 5} \approx 731.62$$

$$AV_{[5,7]} = \frac{186904.57 - 185,444.20}{7 - 5} \approx 730.19$$

$$AV_{[5,6]} = \frac{186172.95 - 185,444.20}{6 - 5} \approx 728.7535$$

The units on each of these quantities are “people per year”.

1.4 The derivative of a function at a point

Motivating 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?

The *instantaneous rate of change* of a function is an idea that sits at the foundation of calculus. It is a generalization of the notion of instantaneous velocity and measures how fast a particular function is changing at a given point. If the original function represents the position of a moving object, this instantaneous rate of change is precisely the velocity of the object. 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 upon increasing a car's velocity by one mile per hour, or the number of dollars added to a mortgage payment for each percentage point increase in interest rate. The instantaneous rate of change can also be interpreted geometrically on the function's graph, and this connection is fundamental to many of the main ideas in calculus.

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

Example 1.4.1. Suppose that f is the function given by the graph below and that a and $a + h$ are the input values as labeled on the x -axis. Use the graph in Figure 1.4.2 to answer the following questions.

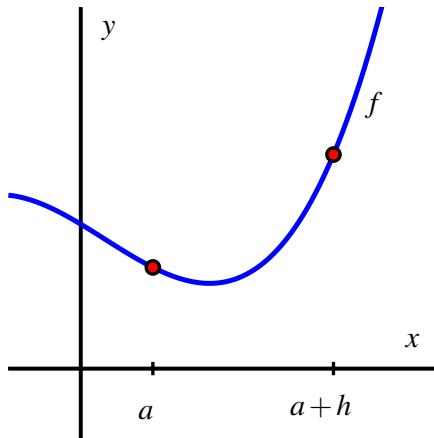


Figure 1.4.2: Plot of $y = f(x)$ for Example 1.4.1.

- Locate and label the points $(a, f(a))$ and $(a + h, f(a + h))$ on the graph.
- 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?
- What is the slope of the line that connects the points $(a, f(a))$ and $(a + h, f(a + h))$?
- 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.

Hint.

- What are the coordinates of the two marked points on the graph of f ?
- Draw the triangle so that one of its legs is horizontal and the other is vertical.
- Remember that the slope of a line segment can be thought of as “rise over run”.
- How does the slope you found compare to the formula for the average rate of change of f on the interval $a, a + h$?

Answer.

- The left marked point on the graph in Figure 1.4.2 is the point $(a, f(a))$, the right marked point on the graph is $(a + h, f(a + h))$.

- b. The lengths of the legs of this triangle are $f(a + h) - f(a)$ and h .
- c. $\frac{f(a+h)-f(a)}{h}$
- d. The average rate of change of the function f on the interval $[a, a + h]$ is the same as the slope of the line segment from $(a, f(a))$ to $(a + h, f(a + h))$.

Solution.

- a. The left marked point on the graph in Figure 1.4.2 is the point $(a, f(a))$, the right marked point on the graph is $(a + h, f(a + h))$.
- b. The length of the vertical leg of this triangle is $f(a + h) - f(a)$, and the length of the horizontal leg of this triangle is $a + h - a = h$.
- c. The slope of the line connecting the two points is $\frac{f(a+h)-f(a)}{h}$, as shown by the triangle drawn in the previous part: the “rise” is the length of the vertical leg, and the “run” is the length of the horizontal leg.
- d. The average rate of change of the function f on the interval $[a, a + h]$ is $AV_{[a,a+h]} = \frac{f(a+h)-f(a)}{h}$, which is exactly what we found in the previous part as the slope of the line segment connecting the points $(a, f(a))$ and $(a + h, f(a + h))$.

1.4.1 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. This instantaneous rate of change of f at a is called “the *derivative* of f at a ,” and is denoted by $f'(a)$.

Derivative at a Point

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

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

provided this limit exists.

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

Note 1.4.3.

- 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 $h \rightarrow 0$. This limit depends on both the function f and the point $x = a$. Since this limit may not exist, not every function has a derivative at every point.

- We say that a function is *differentiable* at $x = a$ if it has a derivative at $x = a$.
- The derivative is a generalization of the instantaneous velocity of a position function: if $y = s(t)$ is a position function of a moving object, $s'(a)$ tells us the instantaneous velocity of the object at time $t = a$.
- Because the units of $\frac{f(a+h)-f(a)}{h}$ are “units of $f(x)$ per unit of x ,” the derivative has these 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.
- 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 actually taking the limit of a collection of slopes of lines. Thus, the derivative itself represents the slope of a particularly important line.

We first consider the derivative at a given value as the slope of a certain line at that value.

When we compute an instantaneous rate of change, we allow the interval $[a, a + h]$ to shrink as $h \rightarrow 0$. We can think of one endpoint of the interval as “sliding towards” the other. In particular, provided that f has a derivative at $(a, f(a))$, the point $(a + h, f(a + h))$ will approach $(a, f(a))$ as $h \rightarrow 0$. Because the process of taking a limit is a dynamic one, it can be helpful to use computing technology to visualize it. One option is a Java applet in which the user is able to control the point that is moving. For a helpful collection of examples, consider the work of David Austin of Grand Valley State University, and this particularly relevant example. For applets that have been built in Geogebra¹, see Marc Renault’s library via Shippensburg University, with this example being especially fitting for our work in this section.

Figure 1.4.4 below shows a sequence of figures with several different lines through the points $(a, f(a))$ and $(a + h, f(a + h))$, 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 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 $h \rightarrow 0$, the secant lines start to approach a single line that passes through the point $(a, f(a))$. If 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 **tangent line** (shown in the right-most figure in green) to the graph of $y = f(x)$ at the point $(a, f(a))$ has slope $m = f'(a)$.

¹You can even consider building your own examples; the fantastic program Geogebra is available for free download and is easy to learn and use.

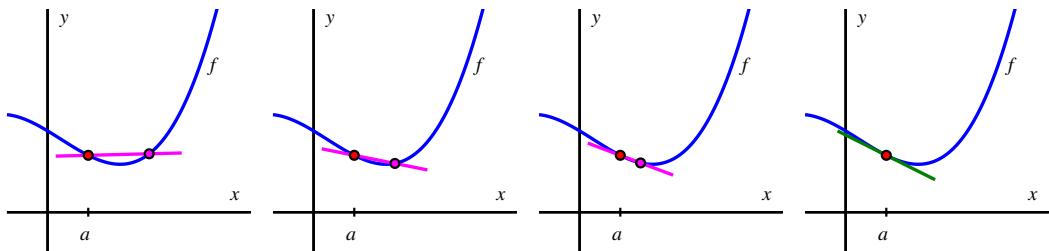


Figure 1.4.4: A sequence of secant lines approaching the tangent line to f at $(a, f(a))$.

If the tangent line at $x = a$ exists, the graph of f looks like a straight line when viewed up close at $(a, f(a))$. In Figure 1.4.5 below, we combine the four graphs from Figure 1.4.4 into a single graph on the left, and zoom in on the box centered at $(a, f(a))$ on the right. 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 $x = a$.

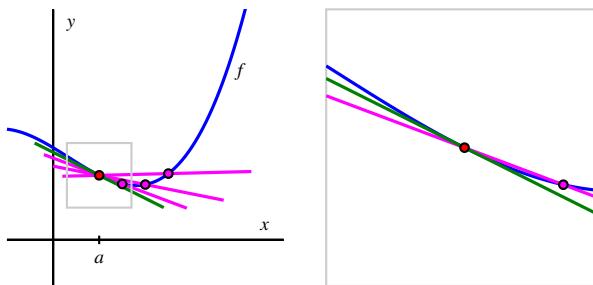


Figure 1.4.5: A sequence of secant lines approaching the tangent line to f at $(a, f(a))$. At right, we zoom in on the point $(a, f(a))$. The slope of the tangent line (in green) to f at $(a, f(a))$ is given by $f'(a)$.

Note 1.4.6. The instantaneous rate of change of f with respect to x at $x = a$, $f'(a)$, 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.

Example 1.4.7 (Using the limit definition of the derivative). For the function $f(x) = x - x^2$, 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.

Hint. The definition of the derivative of f with respect to x evaluated at $x = a$ says

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

How does the value you find for $f'(2)$ manifest in the graph of $f(x) = x - x^2$, particularly near the point $(2, f(2))$?

Answer.

$$f'(2) = -3.$$

The slope of the tangent line to f through the point $(2, -2)$ is -3 .

Solution. From the limit definition, we know that

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

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

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

In order to evaluate the limit, we must simplify the quotient. Expanding and distributing in the numerator gives us

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

Combining like terms, we have

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

Next, because $h \neq 0$ within the limit, we may remove a common factor of h in both the numerator and denominator and find that

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

Finally, we are able to take the limit as h approaches 0, and thus conclude that $f'(2) = -3$. We note that $f'(2)$ is the instantaneous rate of change of f at the point $(2, -2)$. It is also the slope of the tangent line to the graph of $y = x - x^2$ at the point $(2, -2)$. Figure 1.4.8 below shows both the function and the line through $(2, -2)$ with slope $m = f'(2) = -3$.

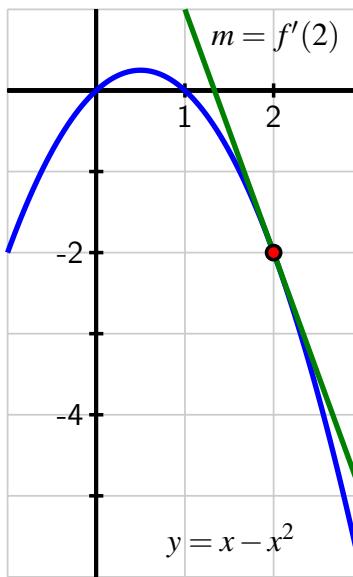


Figure 1.4.8: The tangent line to $y = x - x^2$ at the point $(2, -2)$.

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

Example 1.4.9. Consider the function f whose formula is $f(x) = 3 - 2x$.

- What familiar type of function is f ? What can you say about the slope of f at every value of x ?
- 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?
- 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?
- Without doing any additional computations, what are the values of $f'(2)$, $f'(\pi)$, and $f'(-\sqrt{2})$? Why?

Hint.

- If $f(x) = 3x^2 + 2x - 4$, we say “ f is quadratic.” If $f(x) = 5e^{2x-1}$, we say “ f is exponential.” What do we say about $f(x) = 3 - 2x$?
- Remember that to compute the average rate of change of f on $[a, b]$, we calculate $\frac{f(b) - f(a)}{b - a}$.

- c. Observe that $f(1 + h) = 3 - 2(1 + h) = 3 - 2 - 2h = 1 - 2h$.
- d. Think about the how the graph of f appears. What is the same at every point?

Answer.

- a. f is linear.
- b. The average rate of change on $[1, 4]$, $[3, 7]$, and $[5, 5 + h]$ is -2 .
- c. $f'(1) = -2$.
- d. $f'(2) = -2$, $f'(\pi) = -2$, and $f'(-\sqrt{2}) = -2$, since the slope of a linear function is the same at every point.

Solution.

- a. Because $f(x) = 3 - 2x$ is of the form $f(x) = mx + b$, we call f a *linear* function.
- b. The average rate of change on $[1, 4]$ is $\frac{f(4) - f(1)}{4-1} = \frac{-5-1}{3} = -2$. Similar calculations show the average rate of change on $[3, 7]$ is also -2 . On $[5, 5 + h]$, observe that

$$\begin{aligned}\frac{f(5 + h) - f(5)}{h} &= \frac{(3 - 2(5 + h)) - (3 - 10)}{h} \\ &= \frac{3 - 10 - 2h + 7}{h} \\ &= \frac{-2h}{h} \\ &= -2.\end{aligned}$$

- c. Using the limit definition of the derivative, we find that

$$\begin{aligned}f'(1) &= \lim_{h \rightarrow 0} \frac{f(1 + h) - f(1)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(3 - 2(1 + h)) - (3 - 2)}{h} \\ &= \lim_{h \rightarrow 0} \frac{3 - 2 - 2h - 1}{h} \\ &= \lim_{h \rightarrow 0} \frac{-2h}{h} \\ &= \lim_{h \rightarrow 0} -2 \\ &= -2.\end{aligned}$$

Example 1.4.10. A water balloon is tossed vertically in the air from a window. The balloon's height in feet t seconds after being launched is given by $s(t) = -16t^2 + 16t + 32$. Use this function to respond to each of the following questions.

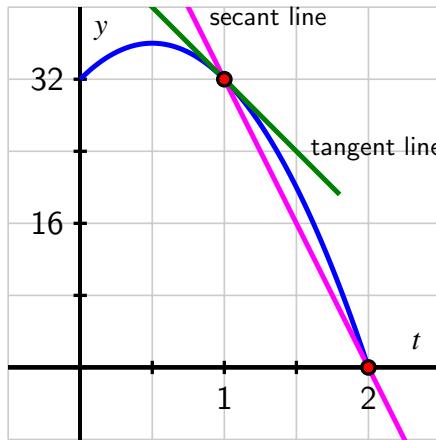
- a. Sketch an accurate, labeled graph of $y = s(t)$ for $t = 0$ to $t = 2$. Label the scale on the axes carefully. You should be able to do this without using computing technology.
- b. Compute the average rate of change of s on the time interval $[1, 2]$. Include units in your answer and write one sentence to explain the meaning of the value you found.
- c. Use the limit definition of the derivative 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 in 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.”

Hint.

- a. Observe that $(t^2 - t - 2) = (t - 2)(t + 1)$ and that $s(t)$ has its vertex at $t = \frac{1}{2}$.
- b. Recall the formula for average rate of change.
- c. Note that $s(1 + h) = -16(1 + h)^2 + 16(1 + h) + 32$.
- d. Think about a secant line and a tangent line.
- e. A line with positive slope is one that is rising; a line with negative slope is one that is falling.

Answer.

- a. The vertex is $(\frac{1}{2}, 36)$.
- b. $\frac{s(2) - s(1)}{2 - 1} = -32$ feet per second.
- c. $s'(1) = -16$.
- d.



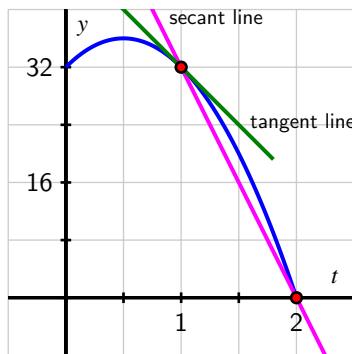
- e. $s'(a)$ is positive whenever $0 \leq a < \frac{1}{2}$; $s'(a)$ is negative whenever $\frac{1}{2} < a < 2$; $s'(\frac{1}{2}) = 0$.

Solution.

- a. Since $s(t) = -16t^2 + 16t + 32 = -16(t^2 - t - 2) = -16(t - 2)(t + 1)$, s has t -intercepts at $(2, 0)$ and $(-1, 0)$; the s -intercept is clearly $(0, 32)$; and the vertex is $(\frac{1}{2}, 36)$. See more in part (d).
- b. Observe that $\frac{s(2) - s(1)}{2 - 1} = \frac{0 - 32}{1} = -32$ feet per second. This value represents the average rate at which the balloon is falling over the time interval from $t = 1$ to $t = 2$.
- c. We compute $s'(1)$ as follows:

$$\begin{aligned}
 s'(1) &= \lim_{h \rightarrow 0} \frac{s(1 + h) - s(1)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(-16(1 + h)^2 + 16(1 + h) + 32) - (-16(1)^2 + 16(1) + 32)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{-16 - 32h - 16h^2 + 16 + 16h + 32 - 32}{h} \\
 &= \lim_{h \rightarrow 0} \frac{-16h - 16h^2}{h} \\
 &= \lim_{h \rightarrow 0} (-16 - 16h) \\
 &= -16.
 \end{aligned}$$

- d. We plot and label the secant line through $(1, s(1))$ and $(2, s(2))$, as well as the tangent line through $(1, s(1))$ with slope $s'(1)$.



- e. Observe that whenever the balloon is rising, its position function is rising, and thus the slope of its tangent line at any such point will be positive. This means that we should find $s'(a)$ to be positive whenever $0 \leq a < \frac{1}{2}$, and similarly $s'(a)$ to be negative whenever $\frac{1}{2} < a < 2$ (which is when the balloon is falling). At the instant $a = \frac{1}{2}$, the balloon is at its vertex and is neither rising nor falling, and at that point, $s'(\frac{1}{2}) = 0$.

Example 1.4.11. 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.

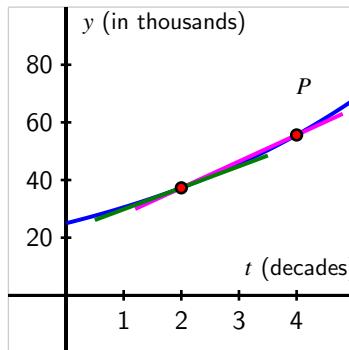
- Sketch an accurate graph of $y = P(t)$ for $t = 0$ to $t = 5$. Label the scale on the axes carefully.
- Compute the average rate of change of P between 2030 and 2050. Include units in your answer and write one sentence to explain (in everyday language) the meaning of the value you found.
- Use the limit definition of the derivative 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.
- 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 in your answer, and write one sentence (using everyday language) to explain the meaning of the value you found.
- 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 $a = 2$.
- 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 ?

Hint.

- $P(t)$ is the standard exponential function, scaled by 25000.
- Use the formula for the average rate of change of a function.
- Because of the exponential nature of $P(t)$, we're not able to simplify $\frac{P(2+h)-P(2)}{h}$ in a way that removes h from the denominator.
- Try using $h = 0.001, 0.0001, 0.00001$ and $h = -0.001, -0.0001, -0.00001$. Be careful not to round or use computing precision that is too limited.
- For the first line, think about the points $(2, P(2))$ and $(4, P(4))$.
- Visualize the slope of the tangent line and how it changes as a point moves along the curve in the positive direction.

Answer.

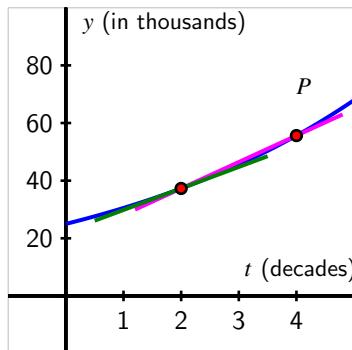
a.



- $AV_{[2,4]} \approx 9171$ people per decade is expected to be the average rate of change of the city's population over the two decades from 2030 to 2050.
- $$P'(2) = \lim_{h \rightarrow 0} \frac{P(2+h) - P(2)}{h} \approx 7458.5,$$
which is measured in people per decade.
- See the graph provided in (a) above. The magenta line has slope equal to the average rate of change of P on $[2, 4]$, while the green line is the tangent line at $(2, P(2))$ with slope $P'(2)$.
- It appears that the tangent line's slope at the point $(a, P(a))$ will increase as a increases.

Solution.

a.



- b. $AV_{[2,4]} = \frac{P(4) - P(2)}{4 - 2} = \frac{25000e^{4/5} - 25000e^{2/5}}{2} \approx 9171$ people per decade is expected to be the average rate of change of the city's population over the two decades from 2030 to 2050.
- c. Note that

$$\begin{aligned} P'(2) &= \lim_{h \rightarrow 0} \frac{P(2+h) - P(2)}{h} = \lim_{h \rightarrow 0} \frac{25000e^{(2+h)/5} - 25000e^{2/5}}{h} \\ &= \lim_{h \rightarrow 0} \frac{25000e^{2/5}e^{h/5} - 25000e^{2/5}}{h} = \lim_{h \rightarrow 0} 25000e^{2/5} \left(\frac{e^{h/5} - 1}{h} \right) \end{aligned}$$

Because there is no way to remove a factor of h from the numerator, we cannot eliminate the h that is making the denominator go to zero, so it appears we need to be content estimating the limit with small values of h .

- d. Using $h = 0.00001$, we find $\frac{P(2+0.00001) - P(2)}{0.00001} \approx 7457$; using $h = -0.00001$, we find $\frac{P(2-0.00001) - P(2)}{-0.00001} \approx 7460$. Averaging these two results, we find that

$$P'(2) = \lim_{h \rightarrow 0} \frac{P(2+h) - P(2)}{h} \approx 7458.5$$

which is measured in people per decade.

- e. See the graph provided in (a) above. The magenta line has slope equal to the average rate of change of P on $[2, 4]$, while the green line is the tangent line at $(2, P(2))$ with slope $P'(2)$.
- f. If we consider the point where $t = a$ and let a start at 0 and then increase, it appears that the tangent line's slope at the point $(a, P(a))$ will increase as a increases.

1.4.2 Summary

- 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 $f(x)$ 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 of a ”) and is defined by the formula

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

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 approaches 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” at the point $(a, f(a))$.
- Limits allow us to move from the rate of change over an interval to the rate of change at a single point.

1.4.3 Exercises

1. (Estimating derivative values graphically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Tangent line to a curve) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Interpreting values and slopes from a graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Estimating a derivative value graphically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Estimating a derivative from the limit definition) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. Consider the graph of $y = f(x)$ provided in Figure 1.4.12.

- a. On the graph of $y = f(x)$, sketch and label the following quantities:
- the secant line to $y = f(x)$ on the interval $[-3, -1]$ and the secant line to $y = f(x)$ on the interval $[0, 2]$.
 - the tangent line to $y = f(x)$ at $x = -3$ and the tangent line to $y = f(x)$ at $x = 0$.
- b. What is the approximate value of the average rate of change of f on $[-3, -1]$? On $[0, 2]$? How are these values related to your work in (a)?
- c. What is the approximate value of the instantaneous rate of change of f at $x = -3$? At $x = 0$? How are these values related to your work in (a)?

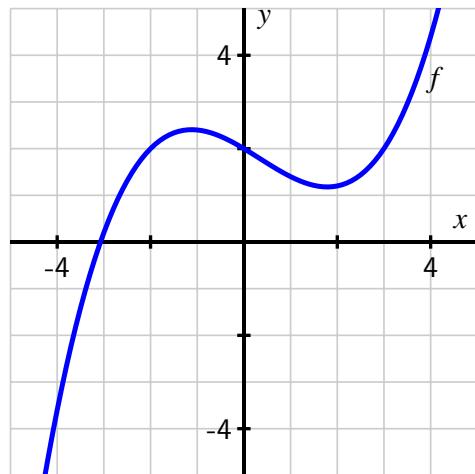
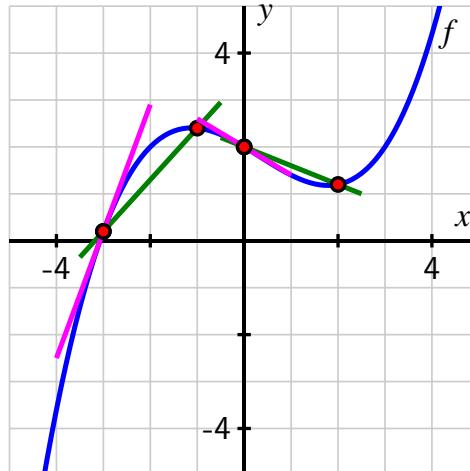


Figure 1.4.12: Plot of $y = f(x)$.

Answer.

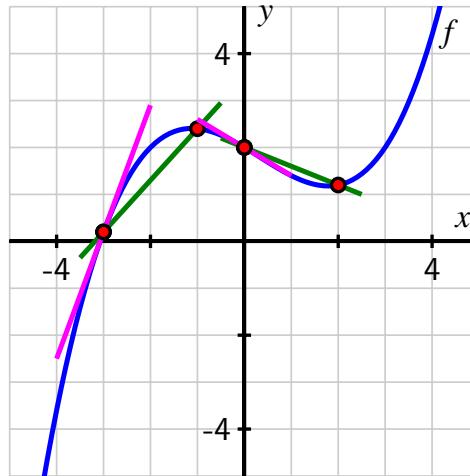
a.



- b. $AV_{[-3, -1]} \approx 1.15$; $AV_{[0, 2]} \approx -0.4$.
- c. $f'(-3) \approx 3$; $f'(0) \approx -\frac{1}{2}$.

Solution.

- a. The requested secant lines are drawn in green, with the noted tangent lines in magenta.



- b. First, $AV_{[-3, -1]} = \frac{f(-1) - f(-3)}{-1 - (-3)} \approx \frac{2.3}{2} = 1.15$. Similarly, $AV_{[0, 2]} = \frac{f(2) - f(0)}{2 - 0} \approx \frac{-0.8}{2} = -0.4$. These values are the respective slopes of the green secant lines plotted in (a).
- c. $f'(-3) \approx 3$ and $f'(0) \approx -\frac{1}{2}$, as these are estimates of the slopes of the tangent lines plotted in magenta.
7. For each of the following prompts, sketch a graph on the provided axes in Figure 1.4.13 of a function that has the stated properties.

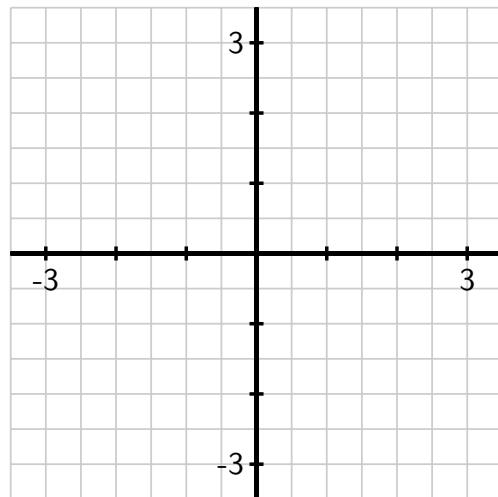
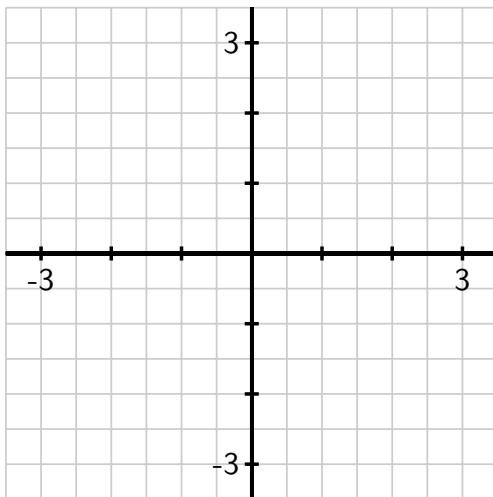


Figure 1.4.13: Axes for plotting $y = f(x)$ in (a) and $y = g(x)$ in (b).

- a. $y = f(x)$ such that

- the average rate of change of f on $[-3, 0]$ is -2 and the average rate of change of f on $[1, 3]$ is 0.5 , and

- the instantaneous rate of change of f at $x = -1$ is -1 and the instantaneous rate of change of f at $x = 2$ is 1 .
- b. $y = g(x)$ such that
- $\frac{g(3)-g(-2)}{5} = 0$ and $\frac{g(1)-g(-1)}{2} = -1$, and
 - $g'(2) = 1$ and $g'(-1) = 0$

Answer.

- a. For instance, you could let $f(-3) = 3$ and have f pass through the points $(-3, 3)$, $(-1, -2)$, $(0, -3)$, $(1, -2)$, and $(3, -1)$ and draw the desired tangent lines accordingly.
- b. For instance, you could draw a function g that passes through the points $(-2, 3)$, $(-1, 2)$, $(1, 0)$, $(2, 0)$, and $(3, 3)$ in such a way that the tangent line at $(-1, 2)$ is horizontal and the tangent line at $(2, 0)$ has slope 1 .

Solution.

- a. For this problem, you should draw a graph so that the slope of the secant line from $(-3, f(-3))$ to $(0, f(0))$ is -2 , and similarly the slope of the secant line from $(1, f(1))$ to $(3, f(3))$ is 0.5 , while the slopes of the tangent lines at $x = -1$ and $x = 2$ are -1 and 1 , respectively. For instance, you could let $f(-3) = 3$ and have f pass through the points $(-3, 3)$, $(-1, -2)$, $(0, -3)$, $(1, -2)$, and $(3, -1)$ and draw the desired tangent lines accordingly.
- b. For this problem, you should draw a graph so that the slope of the secant line from $(-2, g(-2))$ to $(3, f(3))$ is 0 , and similarly the slope of the secant line from $(-1, g(-1))$ to $(1, g(1))$ is -1 , while the slopes of the tangent lines to $y = g(x)$ at $x = -1$ and $x = 2$ are 0 and 1 , respectively. For instance, you could draw a function g that passes through the points $(-2, 3)$, $(-1, 2)$, $(1, 0)$, $(2, 0)$, and $(3, 3)$ in such a way that the tangent line at $(-1, 2)$ is horizontal and the tangent line at $(2, 0)$ has slope 1 .
8. Suppose that the population, P , of China (in billions) can be approximated by the function $P(t) = 1.15(1.014)^t$ where t is the number of years since the start of 1993.
- a. According to the model, what was the total change in the population of China between January 1, 1993 and January 1, 2000? What will be the average rate of change of the population over this time period? Is this average rate of change greater or less than the instantaneous rate of change of the population on January 1, 2000? Explain and justify, being sure to include proper units on all your answers.
- b. According to the model, what is the average rate of change of the population of China in the ten-year period starting on January 1, 2012?
- c. Write an expression involving limits that, if evaluated, would give the exact instantaneous rate of change of the population on today's date. Then estimate the value of this limit (discuss how you chose to do so) and explain the meaning (including units) of the value you have found.
- d. Find an equation for the tangent line to the function $y = P(t)$ at the point where the t -value is given by today's date.

Answer.

- a. $AV_{[0,7]} = \frac{0.1175}{7} \approx 0.01679$ billion people per year; $P'(7) \approx 0.1762$ billion people per year;
 $P'(7) > AV_{[0,7]}$.

- b. $AV_{[19,29]} \approx 0.02234$ billion people/year.

- c. We will say that today's date is July 1, 2015, which means that $t = 22.5$;

$$P'(22.5) = \lim_{h \rightarrow 0} \frac{115(1.014)^{22.5+h} - 115(1.014)^{22.5}}{h};$$

$P'(22.5) \approx 0.02186$ billions of people per year.

- d. $y - 1.57236 = 0.02186(t - 22.5)$.

Solution.

- a. First, $P(7) - P(0) = 115(1.014)^7 - 115(1.014)^0 \approx 0.1175$ billion people is the total population change. It follows that the average rate of change over this time period is $AV_{[0,7]} = \frac{0.1175}{7} \approx 0.01679$ billion people per year. The instantaneous rate of change of the population at $t = 7$ is

$$P'(7) = \lim_{h \rightarrow 0} AV_{[7,7+h]} = \lim_{h \rightarrow 0} \frac{115(1.014)^{7+h} - 115(1.014)^7}{h},$$

and approximating this quantity with small values of h , it follows that $P'(7) \approx 0.1762$. Hence, the instantaneous rate of change at $t = 7$ is greater than the average rate of change of P on the interval $[0, 7]$.

- b. Using the formula for average rate of change,

$$AV_{[19,29]} = \frac{1.15(1.014)^{29} - 1.15(1.014)^{19}}{29 - 19} \approx 0.02234$$

billion people/year.

- c. We will say that today's date is July 1, 2015, which means that $t = 22.5$, so we want to compute $P'(22.5)$. The limit definition tells us that

$$P'(22.5) = \lim_{h \rightarrow 0} \frac{115(1.014)^{22.5+h} - 115(1.014)^{22.5}}{h}.$$

Estimating this limit by using small values of h , we determine that $P'(22.5) \approx 0.02186$, which is measured in billions of people per year.

- d. We want the equation of the tangent line that has slope $P'(22.5) \approx 0.02186$ and that passes through the point $(22.5, P(22.5))$. Using point-slope form, it follows that the line is given by

$$y - 1.57236 = 0.02186(t - 22.5).$$

9. The goal of this problem is to compute the value of the derivative at a point for several different functions, where for each one we do so in three different ways, and then to compare the results to see that each produces the same value.

For each of the following functions, use the limit definition of the derivative to compute the value of $f'(a)$ using three different approaches: strive to use the algebraic approach first (to compute the limit exactly), then test your result using numerical evidence (with small values of h), and finally plot the graph of $y = f(x)$ near $(a, f(a))$ along with the appropriate tangent line to estimate the value of $f'(a)$ visually. Compare your findings among all three approaches; if you are unable to complete the algebraic approach, still work numerically and graphically.

- | | |
|--------------------------------|--|
| a. $f(x) = x^2 - 3x, a = 2$ | d. $f(x) = 2 - x - 1 , a = 1$ |
| b. $f(x) = \frac{1}{x}, a = 1$ | |
| c. $f(x) = \sqrt{x}, a = 1$ | e. $f(x) = \sin(x), a = \frac{\pi}{2}$ |

Answer.

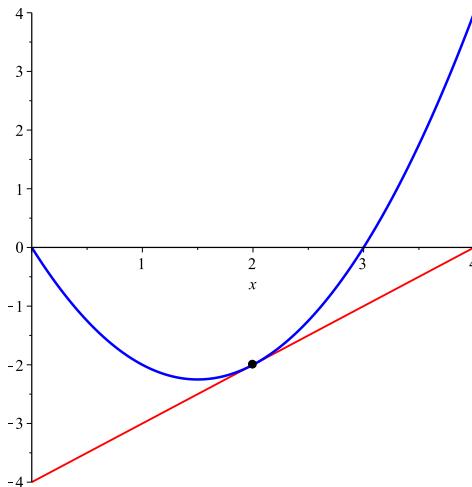
- a. All three approaches show that $f'(2) = 1$.
- b. All three approaches show that $f'(1) = -1$.
- c. All three approaches show that $f'(1) = \frac{1}{2}$.
- d. All three approaches show that $f'(1)$ does not exist.
- e. The first two approaches show that $f'(\frac{\pi}{2}) = 0$.

Solution.

- a. The table below gives values of the difference quotient $\frac{f(a+h)-f(a)}{h}$ for $a = 2$ and several small values of h . This table indicates that $f'(2)$ is approximately 1.

h	-0.0001	-0.001	-0.01	0.01	0.001	0.0001
$\frac{f(2+h)-f(2)}{h}$	0.9999	0.9990	0.9900	1.0100	1.0010	1.0001

The tangent line to f at the point $(2, f(2))$ as shown in the following appears to have a slope that is approximately 1.



Finally, we verify these approximations using an algebraic approach. The definition of the derivative of f at $x = a$ shows that

$$\begin{aligned}
 f'(2) &= \lim_{h \rightarrow 0} \frac{f(2+h) - f(2)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{[(2+h)^2 - 3(2+h)] - (-2)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{[4 + 4h + h^2 - 6 - 3h] + 2}{h}
 \end{aligned}$$

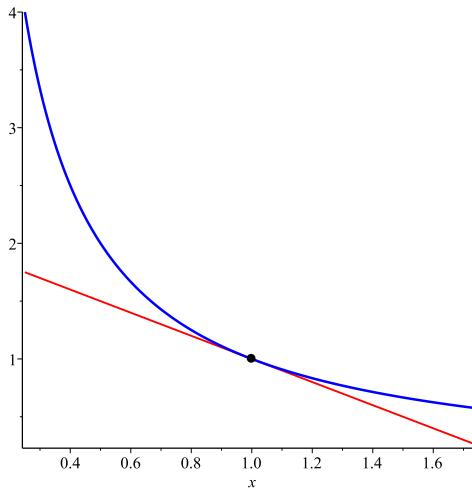
$$\begin{aligned}
 &= \lim_{h \rightarrow 0} \frac{h^2 + h}{h} \\
 &= \lim_{h \rightarrow 0} \frac{h(h + 1)}{h} \\
 &= \lim_{h \rightarrow 0} h + 1 \\
 &= 1
 \end{aligned}$$

This confirms our conclusions using the numeric and geometric approaches.

- b. The following table gives values of the difference quotient $\frac{f(a+h)-f(a)}{h}$ for $a = 1$ and several small values of h . This table indicates that $f'(1)$ is approximately -1 .

h	-0.0001	-0.001	-0.01	0.01	0.001	0.0001
$\frac{f(1+h)-f(1)}{h}$	-1.00010	-1.001001	-1.010101	-0.990099	-0.999001	-0.999900

The tangent line to f at the point $(1, f(1))$ as shown in the following figure appears to have a slope that is approximately -1 .



Finally, we verify these approximations using an algebraic approach. The definition of the derivative of f at $x = a$ shows that

$$\begin{aligned}
 f'(1) &= \lim_{h \rightarrow 0} \frac{f(1+h) - f(1)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\frac{1}{1+h} - 1}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\frac{1}{1+h} - 1}{h} \left(\frac{1+h}{1+h} \right) \\
 &= \lim_{h \rightarrow 0} \frac{1 - (1+h)}{h(1+h)} \\
 &= \lim_{h \rightarrow 0} \frac{-h}{h(1+h)}
 \end{aligned}$$

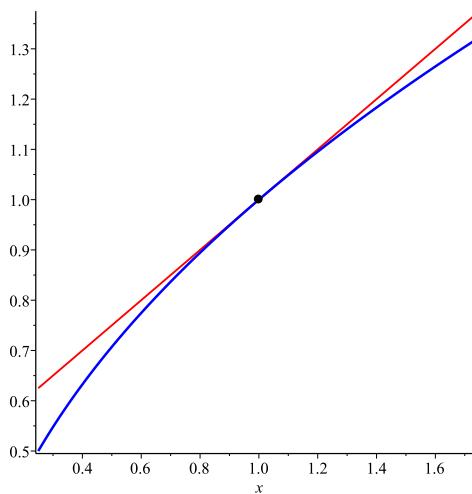
$$= \lim_{h \rightarrow 0} -\frac{1}{1+h} \\ = -1.$$

This confirms our conclusions using the numeric and geometric approaches.

- c. The table below gives values of the difference quotient $\frac{f(a+h)-f(a)}{h}$ for $a = 1$ and several small values of h . This table indicates that $f'(1)$ is approximately 0.5.

h	-0.0001	-0.001	-0.01	0.01	0.001	0.0001
$\frac{f(1+h)-f(1)}{h}$	0.50001300	0.50012510	0.50125629	0.49875620	0.49987500	0.49999000

The tangent line to f at the point $(1, f(1))$ as shown in the figure below appears to have a slope that is approximately 0.5.



Finally, we verify these approximations using an algebraic approach. The definition of the derivative of f at $x = a$ shows that

$$\begin{aligned} f'(1) &= \lim_{h \rightarrow 0} \frac{f(1+h) - f(1)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{1+h} - \sqrt{1}}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{1+h} - 1}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{1+h} - 1}{h} \left(\frac{\sqrt{1+h} + 1}{\sqrt{1+h} + 1} \right) \\ &= \lim_{h \rightarrow 0} \frac{(1+h) - 1}{h(\sqrt{1+h} + 1)} \\ &= \lim_{h \rightarrow 0} \frac{h}{h(\sqrt{1+h} + 1)} \\ &= \frac{1}{\sqrt{1+h} + 1} \end{aligned}$$

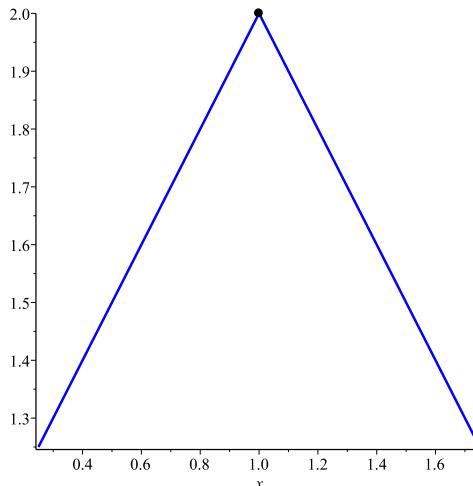
$$= \frac{1}{2}.$$

This confirms our conclusions using the numeric and geometric approaches.

- d. The data in the following table indicates that $\frac{f(1+h)-f(1)}{h}$ has values of 1 for $h < 0$ and values of -1 for $h > 0$. This implies that $f'(1)$ does not exist.

h	-0.0001	-0.001	-0.01	0.01	0.001	0.0001
$\frac{f(1+h)-f(1)}{h}$	1	1	1	-1	-1	-1

The graph of f in the figure below appears to show a sharp corner at the point $(1, f(1))$, also implying that there is no tangent line to f at $(1, f(1))$.



Now we verify these approximations using an algebraic approach. Notice first that

$$\begin{aligned} \frac{f(1+h)-f(1)}{h} &= \frac{[2-|(1+h)-1|]-2}{h} \\ &= \frac{(2-|h|)-2}{h} \\ &= \frac{-|h|}{h} \end{aligned}$$

The definition of the absolute value function tells us that $|h| = h$ whenever $h \geq 0$, while $|h| = -h$ whenever $h < 0$. Thus, for $h > 0$, we have

$$\frac{f(1+h)-f(1)}{h} = \frac{-|h|}{h} = -\frac{h}{h} = -1$$

and for $h < 0$, we have

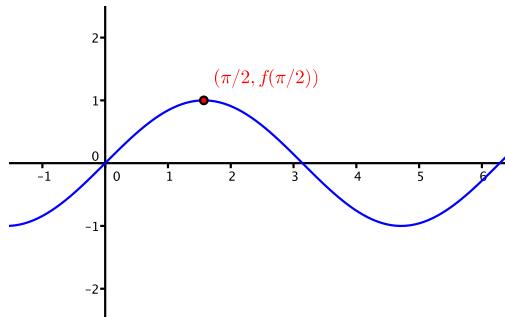
$$\frac{f(1+h)-f(1)}{h} = \frac{-|h|}{h} = -\frac{-h}{h} = 1.$$

This confirms that $\lim_{h \rightarrow 0} \frac{f(1+h)-f(1)}{h}$ does not exist.

- e. The data in the following table indicate that $\frac{f(\frac{\pi}{2}+h)-f(\frac{\pi}{2})}{h}$ approaches 0 as $h \rightarrow 0$.

h	-0.0001	-0.001	-0.01	0.01	0.001	0.0001
$\frac{f(1+h)-f(1)}{h}$	0.00005	0.0005	0.005	-0.005	-0.0005	-0.00005

The graph of f in the figure below shows a horizontal tangent line at the point $(\frac{\pi}{2}, f(\frac{\pi}{2}))$, which also suggests that $f'(\frac{\pi}{2}) = 0$.



Finally, we attempt to verify these approximations using an algebraic approach. Notice first that

$$\begin{aligned}\frac{f(\frac{\pi}{2}+h)-f(\frac{\pi}{2})}{h} &= \frac{\sin(\frac{\pi}{2}+h)-\sin\frac{\pi}{2}}{h} \\ &= \frac{\sin(\frac{\pi}{2}+h)-1}{h}\end{aligned}$$

Using the sum of two angles identity for the sine function, it follows that

$$\sin(\frac{\pi}{2}+h) = \sin\frac{\pi}{2}\cos(h) + \cos\frac{\pi}{2}\sin(h) = 1 \cdot \cos(h) + 0 \cdot \sin(h) = \cos(h).$$

Hence,

$$\begin{aligned}f'(\frac{\pi}{2}) &= \lim_{h \rightarrow 0} \frac{f(\frac{\pi}{2}+h)-f(\frac{\pi}{2})}{h} \\ &= \lim_{h \rightarrow 0} \frac{\cos(h)-1}{h}\end{aligned}$$

At this point, we encounter a limit that we are unable to evaluate further using algebraic means. Like the earlier limit above, we can estimate the limit using small values of h . Doing so confirms that this limit also approaches 0, and we hence conclude that $f'(\frac{\pi}{2}) = 0$.

1.5 The derivative function

Motivating 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 $f'(x)$ related to the graph of $f(x)$?
- What are some examples of functions f for which f' is undefined at one or more points?

We now know that the instantaneous rate of change of a function $f(x)$ at $x = a$, or equivalently the slope of the tangent line to the graph of $y = f(x)$ at $x = a$, is given by the value $f'(a)$. In all of our examples so far, we have 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 the derivative at a single point to computing a formula for $f'(a)$ at any point a . Indeed, the process of “taking the derivative” generates a new function, denoted by $f'(x)$, derived from the original function $f(x)$.

Example 1.5.1. Consider the function $f(x) = 4x - x^2$.

- Use the limit definition to compute the derivative values: $f'(0)$, $f'(1)$, $f'(2)$, and $f'(3)$.
- 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.)
- 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 .

Hint.

- $$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$
- Is there a pattern with the four values you found in (a)?
- Based on (a) and (b), what type of familiar function might the derivative be? What do formulas for this type of function look like?

Answer.

- a. $f'(0) = 4, f'(1) = 2, f'(2) = 0, f'(3) = -2.$
- b. $f'(4) = -4, f'(5) = -6.$
- c. $f'(a) = -2a + 4.$

Solution.

- a. Since $f(0) = 0$ and $f(0 + h) = f(h) = 4h - h^2$, we can say that

$$\begin{aligned} f'(0) &= \lim_{h \rightarrow 0} \frac{f(0 + h) - f(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(4h - h^2) - 0}{h} \\ &= \lim_{h \rightarrow 0} (4 - h) \\ &= 4. \end{aligned}$$

Since $f(1) = 4(1) - (1)^2 = 3$ and

$$\begin{aligned} f(1 + h) &= 4(1 + h) - (1 + h)^2 \\ &= 4 + 4h - (1 + 2h + h^2) \\ &= 3 + 2h - h^2, \end{aligned}$$

we can say that

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

Similar work shows that $f'(2) = 0$ and $f'(3) = -2$.

- b. In part (a), each time we increased the input value by 1, the derivative value decreased by 2. This leads us to believe that $f'(4) = -4$ and $f'(5) = -6$, assuming the linear pattern continues.
- c. The derivative values seem to fit a linear pattern, with a slope of -2 and a y -intercept of 4 . A reasonable formula for $f'(a)$, then, is $f'(a) = -2a + 4$.

1.5.1 How the derivative is itself a function

In your work in Example 1.5.1 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 $f'(5) = -6$. We also observe that the particular value of a has very little effect on the process of computing the value of the derivative through the limit definition. 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,

$$\begin{aligned} f'(a) &= \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \rightarrow 0} \frac{4(a+h) - (a+h)^2 - (4a - a^2)}{h} \\ &= \lim_{h \rightarrow 0} \frac{4a + 4h - a^2 - 2ha - h^2 - 4a + a^2}{h} = \lim_{h \rightarrow 0} \frac{4h - 2ha - h^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{h(4 - 2a - h)}{h} = \lim_{h \rightarrow 0} (4 - 2a - h). \end{aligned}$$

Here we observe that neither 4 nor $2a$ depend on the value of h , so as h tends to 0, $(4 - 2a - h)$ approaches $(4 - 2a)$. Thus, $f'(a) = 4 - 2a$.

This result is consistent with the specific values we found above: e.g., $f'(3) = 4 - 2(3) = -2$. And indeed, our work confirms that the value of a 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. For the function $f(x) = 4x - x^2$, it follows that $f'(x) = 4 - 2x$.

Because the value of the derivative function is linked to the graph of the original function, it makes sense to look at both of these functions plotted on the same domain.

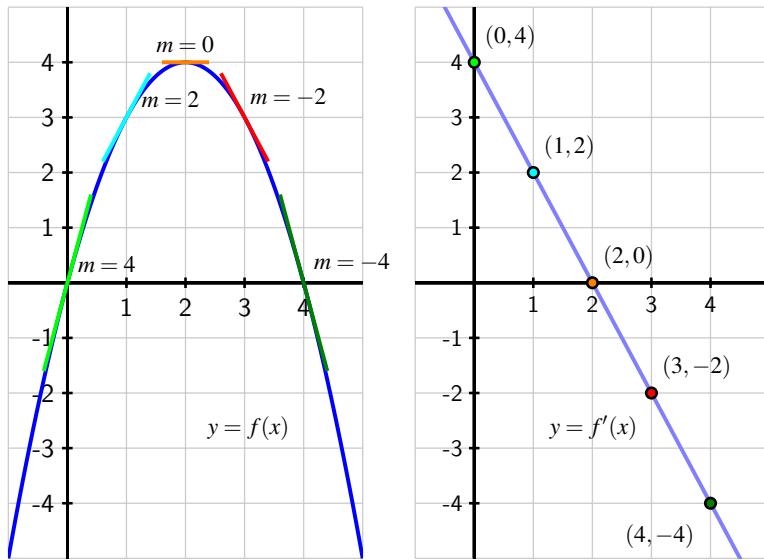


Figure 1.5.2: The graphs of $y = f(x)$ (at left) with $f(x) = 4x - x^2$, and of $y = f'(x)$ (at right), where $f'(x) = 4 - 2x$. Slopes on the graph of $y = f(x)$ correspond to y -coordinates on the graph of $y = f'(x)$.

The left half of Figure 1.5.2 above, shows a plot of $y = f(x)$, where $f(x) = 4x - x^2$, together with a selection of tangent lines at the points we considered in Example 1.5.1. The right half of the figure shows a plot of $y = f'(x)$, where $f'(x) = 4 - 2x$ with emphasis on the y -coordinates of the derivative graph at the same selection of points. Notice the connection between colors in the left and right graphs: 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 lefthand graph is the same as the y -value at the corresponding point on the righthand graph.* That is, at each respective value of x , the slope of the tangent line to the original function is the same as the output of the derivative function. 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 righthand graph of $y = f'(x)$, the units on the vertical axis are units of f per unit of x .

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 Austin and Renault¹.

In Section 1.4 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.

¹David Austin, <http://gvsu.edu/s/5r>; Marc Renault, <http://gvsu.edu/s/5p>.

Derivative Function

Let f be a function and x a value in the function's domain. We define a new function called f' to be the **derivative of f** , where f' is given by the formula

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

provided this limit exists.

We now have two different ways of thinking about the derivative function:

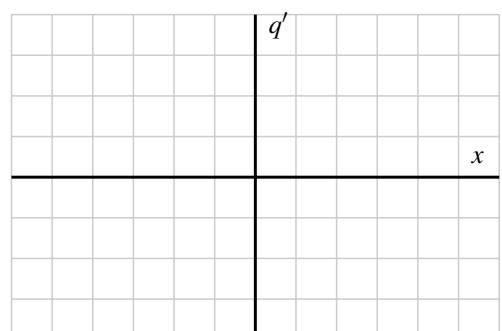
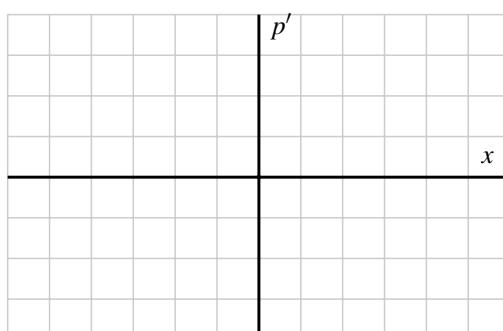
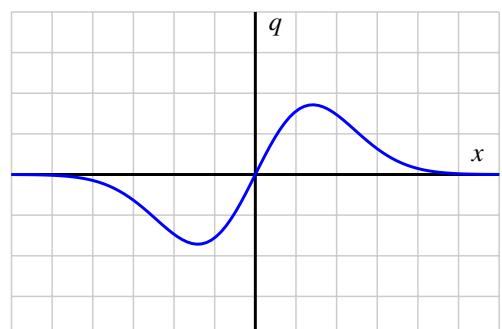
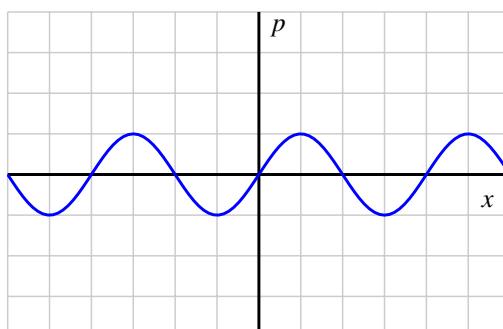
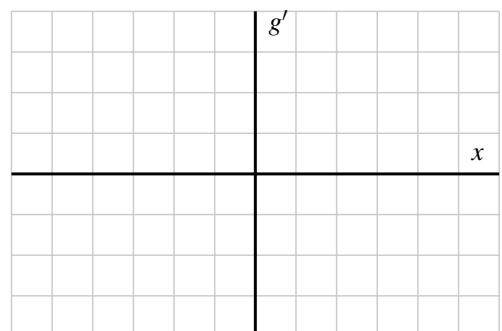
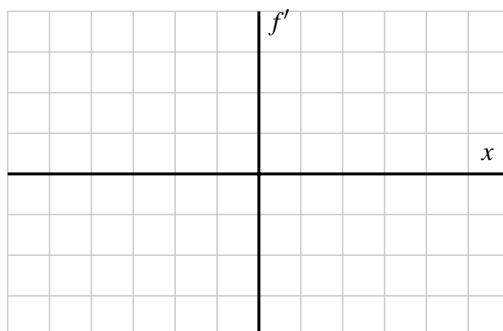
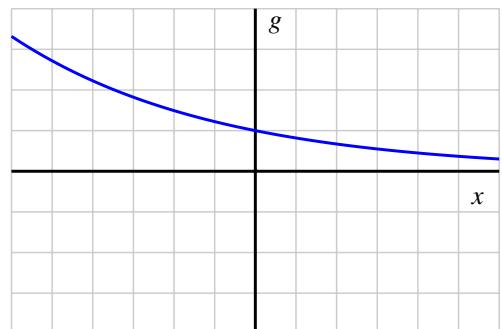
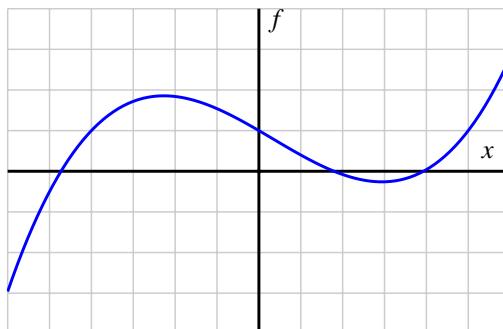
1 given a graph of $y = f(x)$, how does this graph lead to the graph of the derivative function $y = f'(x)$? and

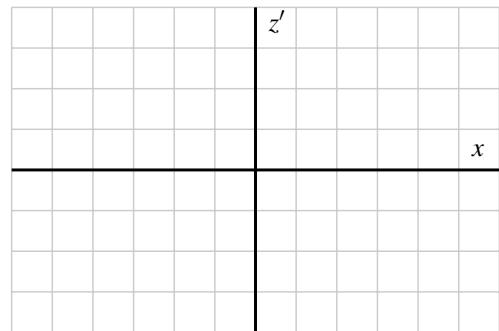
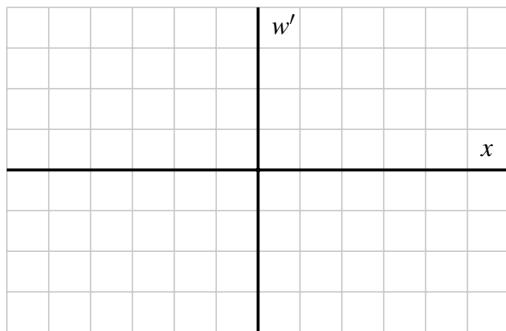
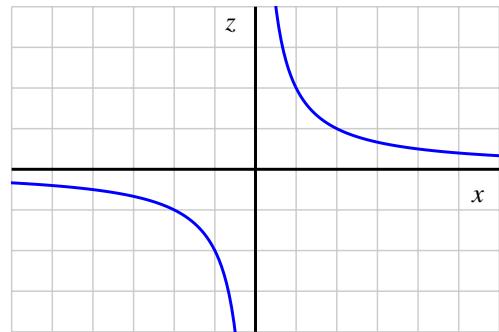
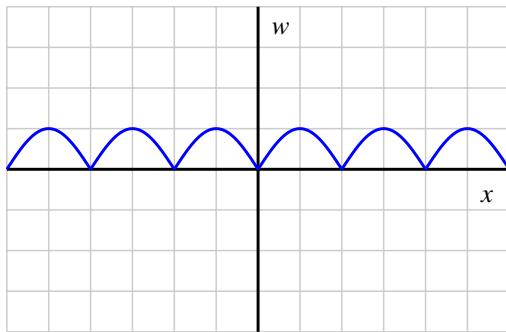
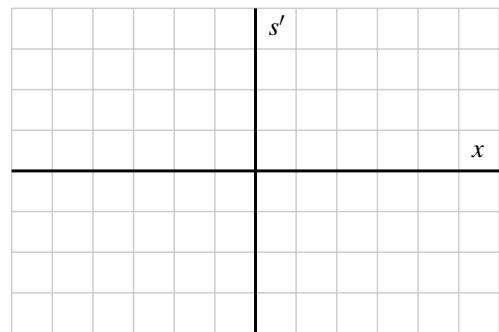
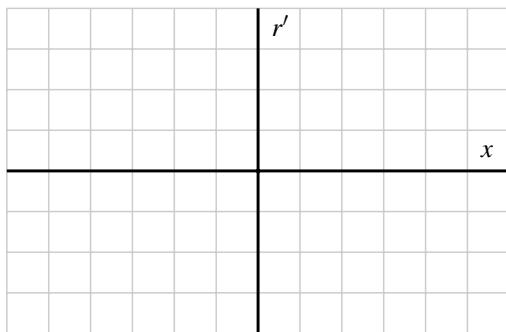
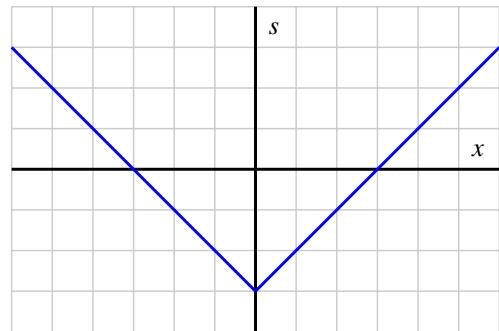
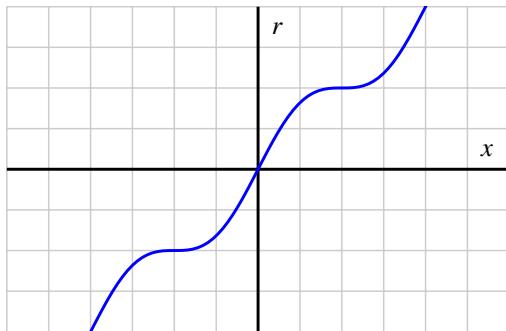
2 given a formula for $y = f(x)$, how does the limit definition of derivative generate a formula for $y = f'(x)$?

Both of these issues are explored in the following examples.

Example 1.5.3. Below you will find the graphs of eight different functions, each on a grid with scale 1×1 . For each given graph, sketch an approximate graph of its derivative function on the axes immediately below. Assume the *horizontal* scale of the grid for each derivative graph is identical to that for the original function. If necessary, adjust the vertical scale on the axes for the graphs of each derivative. Label all axes (vertical and horizontal) with the scales you use.

When you are finished with all eight graphs, write several sentences that describe your overall process for sketching the graph of the derivative function, given the graph of 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?

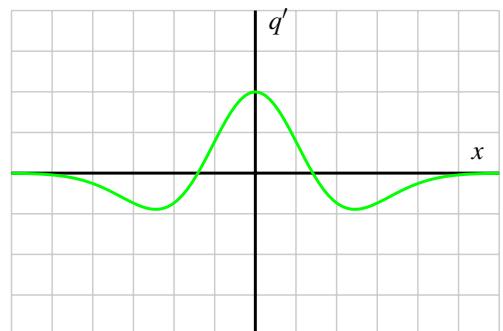
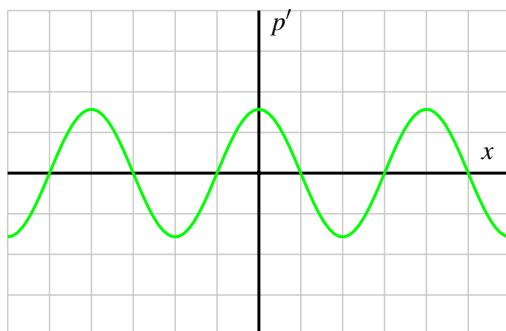
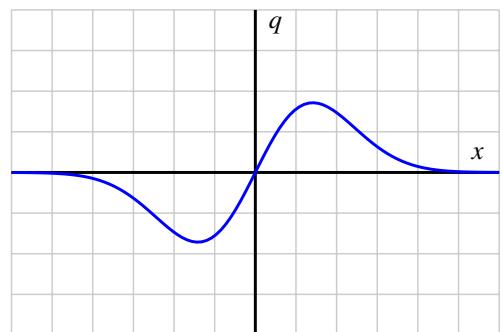
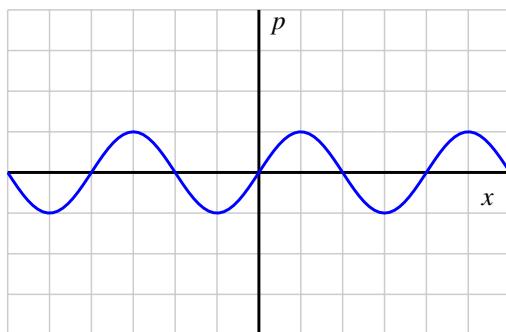
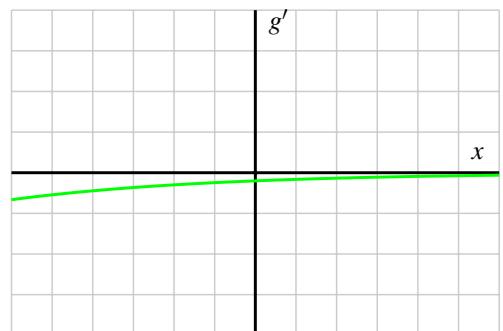
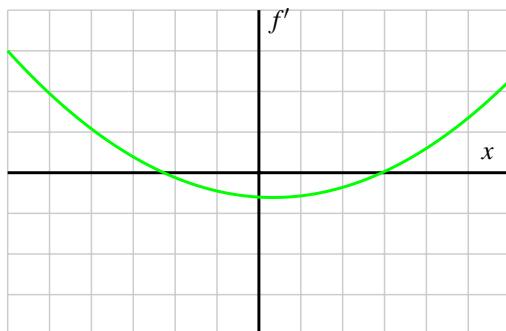
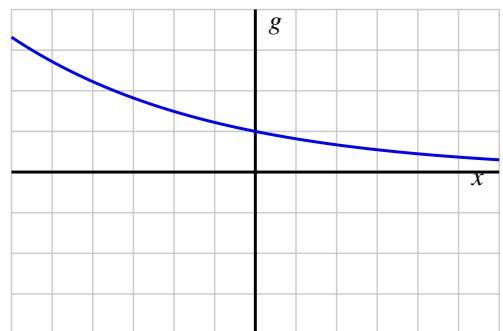
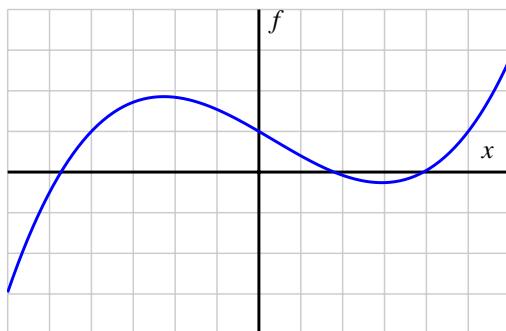


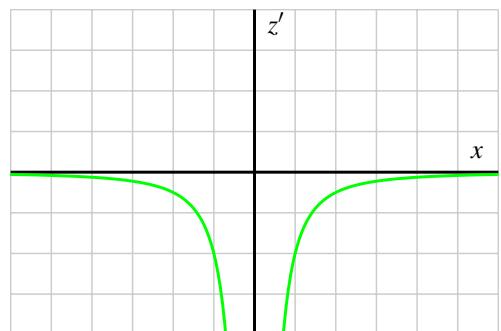
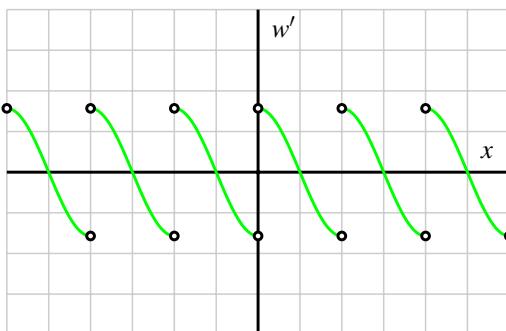
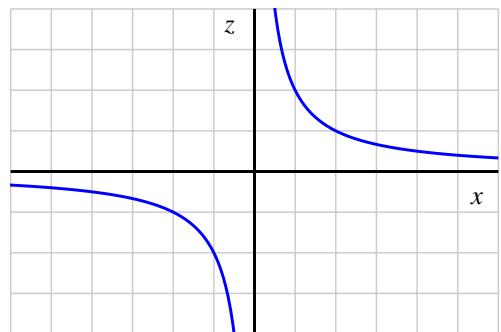
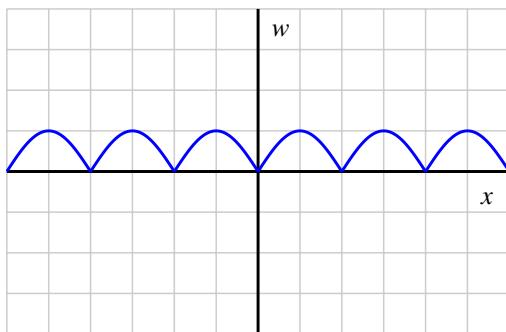
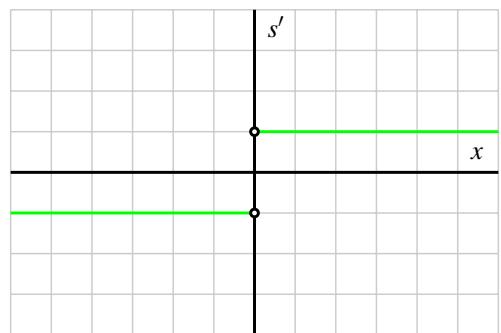
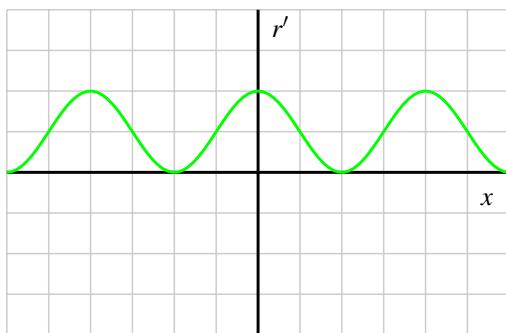
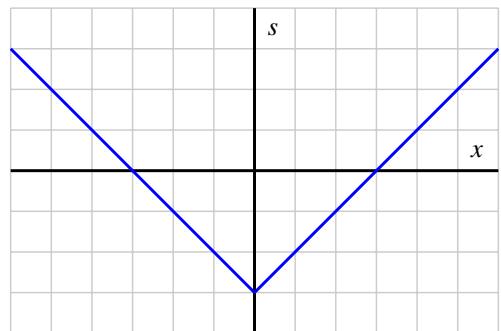
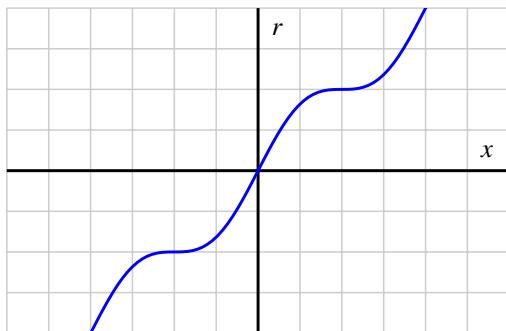


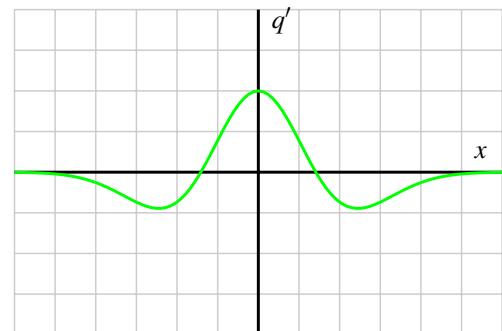
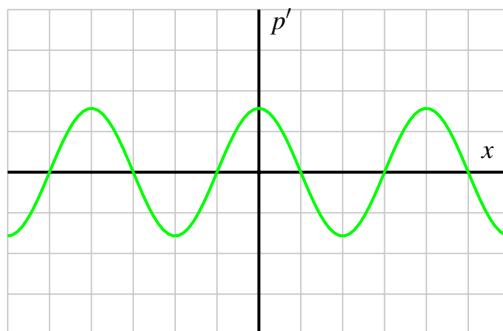
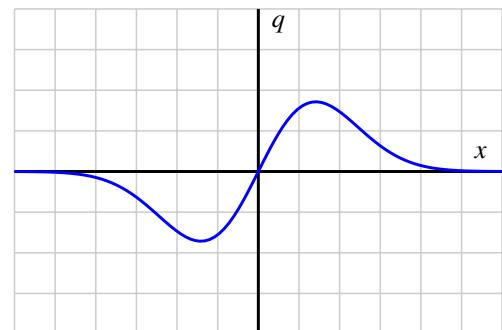
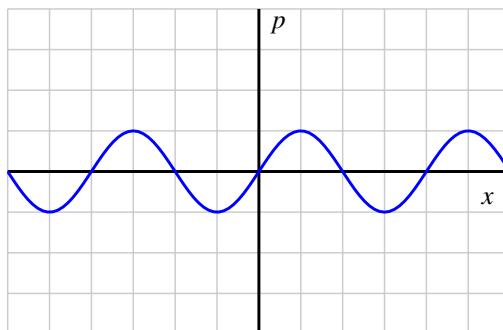
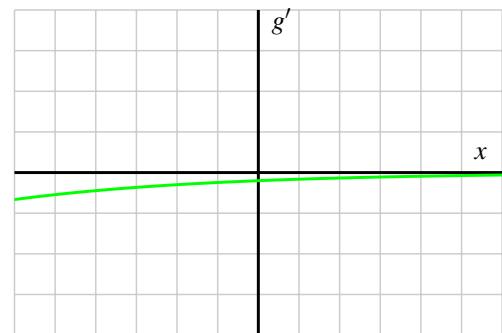
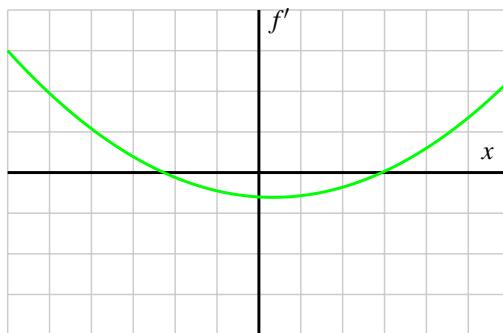
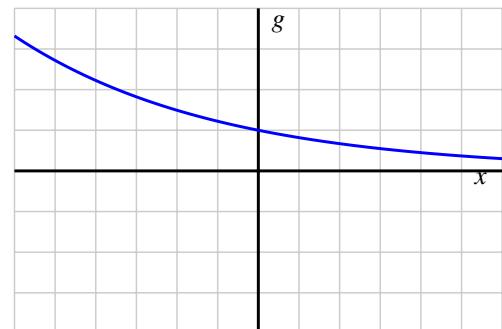
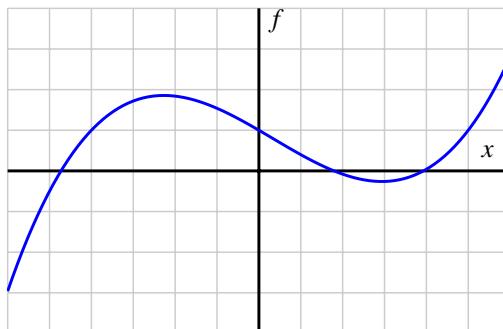
Hint. Points where the slope of the tangent line is equal to zero are particularly important. Try finding these points first in your effort to plot $y = f'(x)$ and plotting those zero values

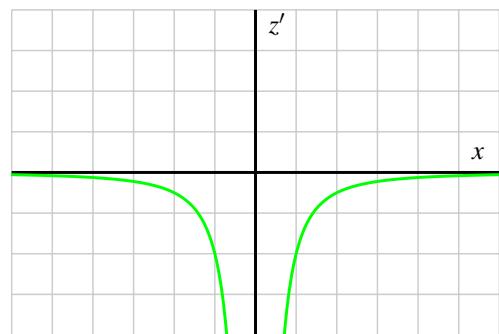
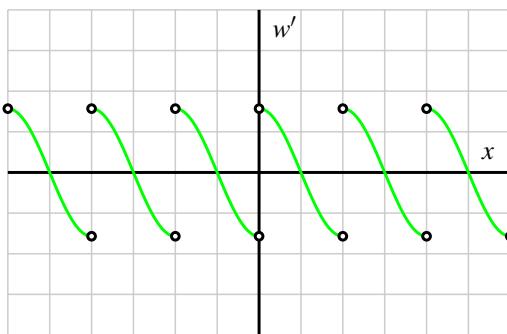
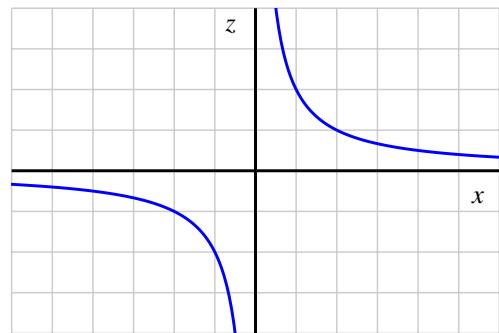
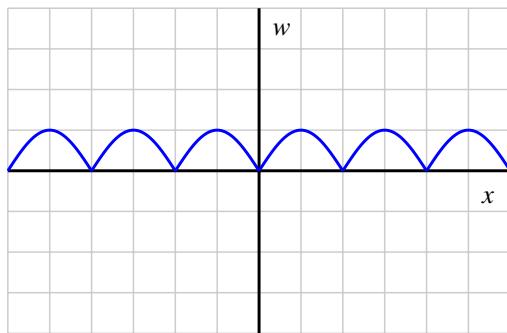
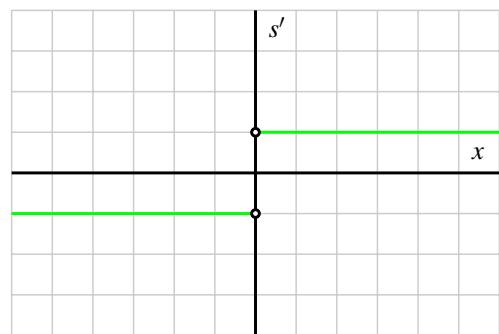
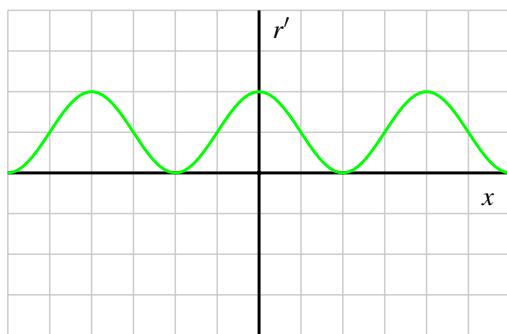
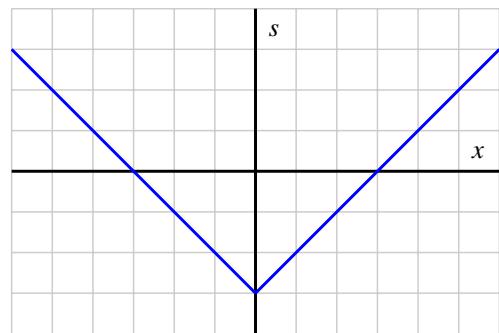
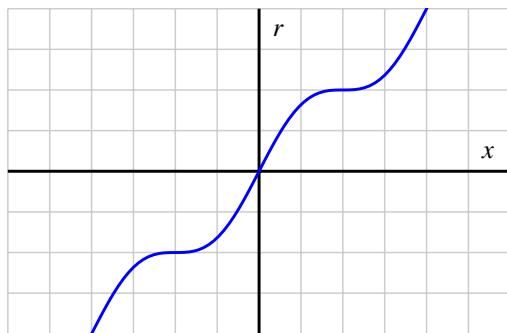
on the axes where you'll graph $y = f'(x)$.

Answer.





Solution.



For a dynamic investigation that allows you to experiment with graphing f' when given the

graph of f , see <http://gvsu.edu/s/8y>.²

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' in Figure 1.5.2. Following Example 1.5.3, we now understand that we could have constructed a fairly accurate graph of $f'(x)$ *without* knowing a formula for either f or f' . Even so, it is useful to know a formula for the derivative function whenever it is possible to find one.

In the next example, we further explore the more algebraic approach to finding a derivative: given a formula for $y = f(x)$, the limit definition of the derivative will be used to develop a formula for $f'(x)$.

Example 1.5.4. 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. $f(x) = 1$

c. $p(z) = z^2$

e. $F(t) = \frac{1}{t}$

b. $g(t) = t$

d. $q(s) = s^3$

f. $G(y) = \sqrt{y}$

Hint.

a. What is the slope of the function at every point?

b. What is the slope of the function at every point?

c. $p(z + h) = (z + h)^2$

d. $q(s + h) = (s + h)^3$

e. $F(t + h) = \frac{1}{t+h}$

f. $G(y + h) = \sqrt{y + h}$

Answer.

a. $f'(x) = 0$.

b. $g'(t) = 1$.

c. $p'(z) = 2z$.

d. $q'(s) = 3s^2$.

e. $F'(t) = \frac{-1}{t^2}$.

²Marc Renault, Calculus Applets Using Geogebra.

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

Solution.

- The instantaneous rate of change of a linear function at a given point is the same, regardless of the point chosen. This is because linear functions have a constant slope, so any tangent line to a linear function coincides with the function itself (i.e., they are the same line). In this case, $f'(x) = 0$ because the slope of the tangent line to the horizontal line given by $f(x) = 1$ is zero at every value of x .
- $g'(t) = 1$ because the slope of the tangent line to the line given by $g(t) = t$ is 1 at every value of t .

c.

$$\begin{aligned}
 p'(z) &= \lim_{h \rightarrow 0} \frac{p(z+h) - p(z)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(z+h)^2 - z^2}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(z^2 + 2zh + h^2) - z^2}{h} \\
 &= \lim_{h \rightarrow 0} \frac{2zh + h^2}{h} \\
 &= \lim_{h \rightarrow 0} (2z + h) \\
 &= 2z
 \end{aligned}$$

d.

$$\begin{aligned}
 q'(s) &= \lim_{h \rightarrow 0} \frac{q(s+h) - q(s)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(s+h)^3 - s^3}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(s^3 + 3s^2h + 3sh^2 + h^3) - s^3}{h} \\
 &= \lim_{h \rightarrow 0} \frac{3s^2h + 3sh^2 + h^2}{h} \\
 &= \lim_{h \rightarrow 0} (3s^2 + 3sh + h) \\
 &= 3s^2
 \end{aligned}$$

e.

$$\begin{aligned}
 F'(t) &= \lim_{h \rightarrow 0} \frac{F(t+h) - F(t)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\left(\frac{1}{t+h} - \frac{1}{t}\right)}{h}
 \end{aligned}$$

$$\begin{aligned}
 &= \lim_{h \rightarrow 0} \frac{\binom{t-(t+h)}{t(t+h)}}{h} \\
 &= \lim_{h \rightarrow 0} \frac{-h}{ht(t+h)} \\
 &= \lim_{h \rightarrow 0} \frac{-1}{t(t+h)} \\
 &= \frac{-1}{t^2}
 \end{aligned}$$

f.

$$\begin{aligned}
 G'(y) &= \lim_{h \rightarrow 0} \frac{G(y+h) - G(y)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\sqrt{y+h} - \sqrt{y}}{h} \\
 &= \lim_{h \rightarrow 0} \left(\frac{\sqrt{y+h} - \sqrt{y}}{h} \right) \cdot \left(\frac{\sqrt{y+h} + \sqrt{y}}{\sqrt{y+h} + \sqrt{y}} \right) \\
 &= \lim_{h \rightarrow 0} \frac{(y+h) - y}{h(\sqrt{y+h} + \sqrt{y})} \\
 &= \lim_{h \rightarrow 0} \frac{h}{h(\sqrt{y+h} + \sqrt{y})} \\
 &= \lim_{h \rightarrow 0} \frac{1}{\sqrt{y+h} + \sqrt{y}} \\
 &= \frac{1}{2\sqrt{y}}
 \end{aligned}$$

1.5.2 Summary

- The limit definition of the derivative, $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$, produces a value for each x at which the derivative is defined, and this leads to a new function $y = f'(x)$. It is especially important to note that taking the derivative is a process that starts with a given function (f) and produces a new, related function (f').
- There is essentially no difference between writing $f'(a)$ (as we did regularly in Section 1.4) 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 *y-coordinates* on the derivative's graph correspond to *slopes* on the original function's graph.
- In Example 1.5.3, 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.

1.5.3 Exercises

1. (The derivative function graphically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (The derivative function graphically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Applying the limit definition of the derivative) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Sketching the derivative) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Comparing function and derivative values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Limit definition of the derivative for a rational function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. 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 in Figure 1.5.5, sketch a possible graph of $y = f(x)$. Explain why your graph meets the stated criteria.
 - b. 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.

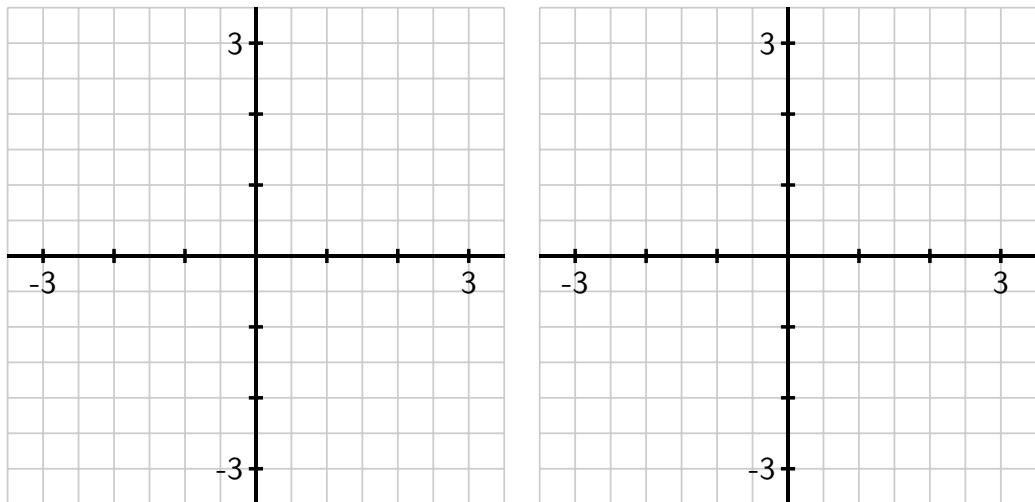
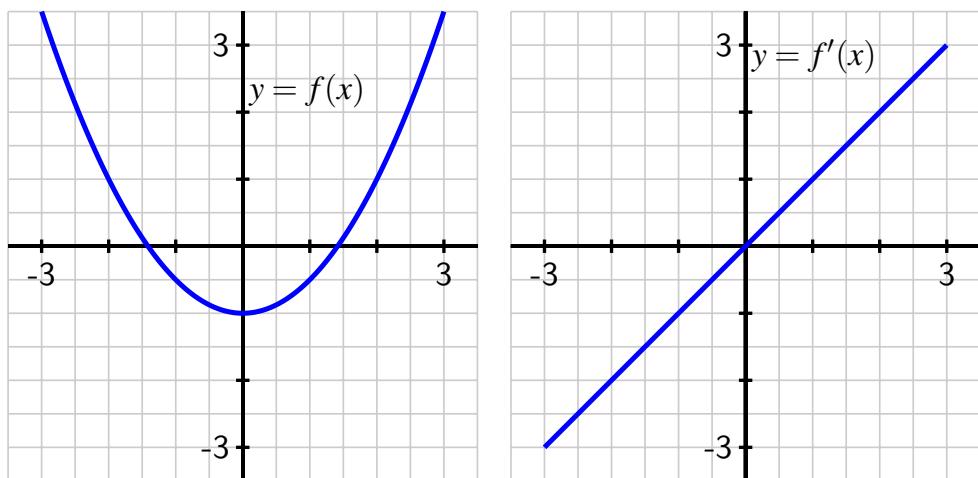


Figure 1.5.5: Axes for plotting $y = f(x)$ in (a) and $y = f'(x)$ in (b).

Answer.

- See the figure below.
- See the figure below.

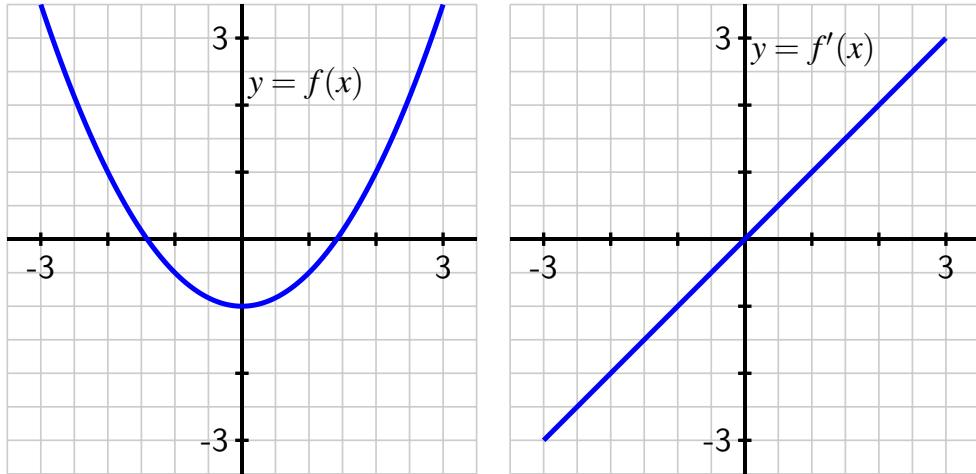


- One example of a formula for f is $f(x) = \frac{1}{2}x^2 - 1$.

Solution.

- The fact that f is differentiable everywhere means that the graph of f is smooth everywhere. The slopes of the tangent lines to f are negative but increasing on the interval $(-\infty, 0)$ and positive and increasing on the interval $(0, \infty)$, with a slope of 0 when $x = 0$. This is the kind of behavior that a quadratic function possesses, so we could guess that f has a graph something like that shown in the figure below.

- b. Because the change in the derivative values is constant, it looks like f' is linear with a slope of 1, passing through the point $(0, 0)$, so it is reasonable to guess that $f'(x) = x$. A plot of f' is shown at right in the figure below.



- c. A natural guess is $f(x) = x^2$; since we need the function to pass through the point $(1, -2)$, we might try $f(x) = x^2 - 3$. Using the limit definition, we have

$$\begin{aligned}
 f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{[(x+h)^2 - 3] - [x^2 - 3]}{h} \\
 &= \lim_{h \rightarrow 0} \frac{[x^2 + 2xh + h^2 - 3] - [x^2 - 3]}{h} \\
 &= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} \\
 &= \lim_{h \rightarrow 0} 2x + h \\
 &= 2x.
 \end{aligned}$$

So this guess is close, but is off by a factor of 2, since we want $f'(x) = x$. Instead, if we use $f(x) = \frac{1}{2}x^2 - 1$ (note that we chose the “-1” so that $f(-2) = 1$, then we have

$$\begin{aligned}
 f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\left[\frac{1}{2}(x+h)^2 - 1\right] - \left[\frac{1}{2}x^2 - 1\right]}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\frac{1}{2}[x^2 + 2xh + h^2] - \frac{1}{2}[x^2]}{h} \\
 &= \lim_{h \rightarrow 0} \frac{1}{2} \frac{2xh + h^2}{h} \\
 &= \lim_{h \rightarrow 0} x + \frac{h}{2}
 \end{aligned}$$

$$= x.$$

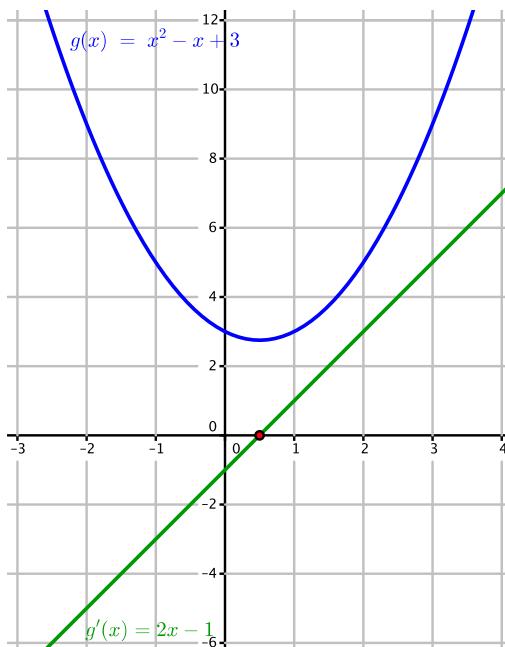
This appears to be the correct function f .

8. Consider the function $g(x) = x^2 - x + 3$.
- Use the limit definition of the derivative to determine a formula for $g'(x)$.
 - Use a graphing utility to plot both $y = g(x)$ and your result for $y = g'(x)$; does your formula for $g'(x)$ generate the graph you expected?
 - Use the limit definition of the derivative to find a formula for $p'(x)$ where $p(x) = 5x^2 - 4x + 12$.
 - Compare and contrast the formulas for $g'(x)$ and $p'(x)$ you have found. How do the constants 5, 4, 12, and 3 affect the results?

Answer.

a. $g'(x) = 2x - 1$.

b.



c. $p'(x) = 10x - 4$.

- d. The constants 3 and 12 don't seem to affect the results at all. The coefficient -4 on the linear term in $p(x)$ appears to make the " -4 " appear in $p'(x) = 10x - 4$. The leading coefficient 5 in $p(x) = 5x^2 - 4x + 12$ leads to the coefficient of " 10 " in $p'(x) = 10x - 4$.

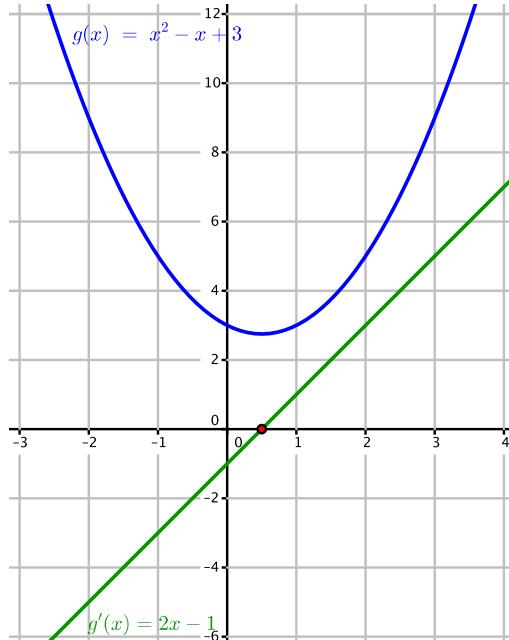
Solution.

- a. By definition,

$$g'(x) = \lim_{h \rightarrow 0} \frac{g(x + h) - g(x)}{h}$$

$$\begin{aligned}
 &= \lim_{h \rightarrow 0} \frac{(x+h)^2 - (x+h) + 3 - (x^2 - x + 3)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x - h + 3 - x^2 + x - 3}{h} \\
 &= \lim_{h \rightarrow 0} \frac{2xh + h^2 - h}{h} \\
 &= \lim_{h \rightarrow 0} \frac{h(2x + h - 1)}{h} \\
 &= \lim_{h \rightarrow 0} (2x + h - 1) \\
 &= 2x - 1.
 \end{aligned}$$

b.



In the above figure, we see plots of both g and g' . We observe that the point $(0.5, 2.75)$ is the vertex of the quadratic function g , and at this point the slope of the tangent line to $g(x)$ is zero. This aligns with the point $(0.5, 0)$ where $y = g'(x)$ crosses the x -axis. In addition, we note that $g'(x)$ is negative for $x < 0.5$, which corresponds to where g is decreasing and has tangent lines with slopes that are negative. Similarly, the values of g' are positive for $x > 0.5$, which align with the values of slopes we see on the original function g .

c. By definition,

$$\begin{aligned}
 p'(x) &= \lim_{h \rightarrow 0} \frac{p(x+h) - p(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{5(x+h)^2 - 4(x+h) + 12 - (5x^2 - 4x + 12)}{h}
 \end{aligned}$$

$$\begin{aligned}
 &= \lim_{h \rightarrow 0} \frac{5x^2 + 10xh + 5h^2 - 4x - 4h + 12 - 5x^2 + 4x - 12}{h} \\
 &= \lim_{h \rightarrow 0} \frac{10xh + 5h^2 - 4h}{h} \\
 &= \lim_{h \rightarrow 0} \frac{h(10x + 5h - 4)}{h} \\
 &= \lim_{h \rightarrow 0} (10x + 5h - 4) \\
 &= 10x - 4.
 \end{aligned}$$

d. For $g(x) = x^2 - x + 3$, we found that $g'(x) = 2x - 1$. For $p(x) = 5x^2 - 4x + 12$, we determined that $p'(x) = 10x - 4$. The constants 3 and 12 don't seem to affect the results at all, and that makes sense because those numbers only serve to shift the graphs of g and p vertically, which does nothing to change the slope. The coefficient -4 on the linear term in $p(x)$ appears to make the " -4 " appear in $p'(x) = 10x - 4$. That, too, makes sense in light of the fact that if we considered only the linear function $L(x) = -4x$, the slope would everywhere be -4 , in contrast to the coefficient -1 found on the linear term in $g(x) = x^2 - x + 3$, which leads to the constant -1 in $g'(x) = 2x - 1$. Finally, the leading coefficient 5 in $p(x) = 5x^2 - 4x + 12$ leads to the coefficient of " 10 " in $p'(x) = 10x - 4$. This makes sense because if we considered only $y = 5x^2$, the 5 would make the graph 5 times as steep as the graph of $y = x^2$, and thus it affects the derivative proportionately.

9. 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 in Figure 1.5.6.

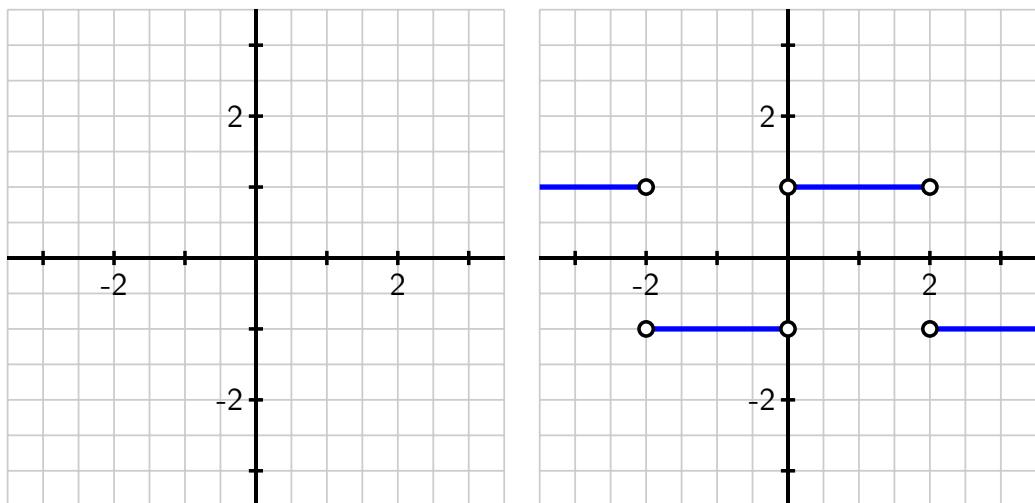


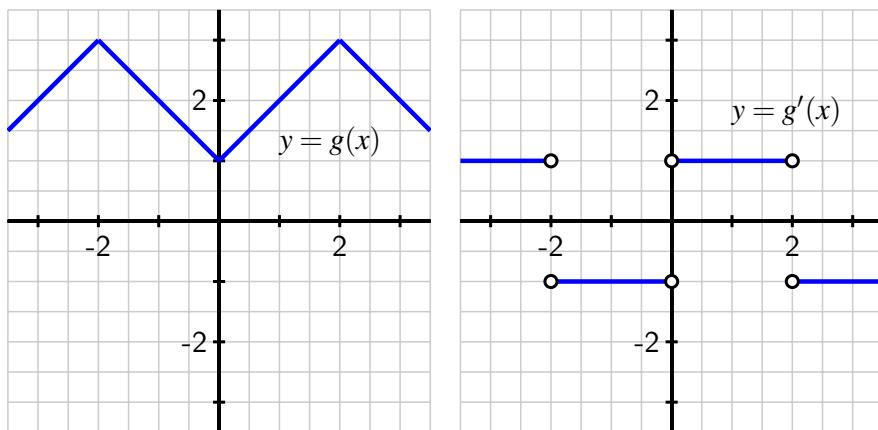
Figure 1.5.6: Axes for plotting $y = g(x)$ and, at right, the graph of $y = g'(x)$.

- 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 in Figure 1.5.6, sketch an accurate graph of $y = g(x)$.

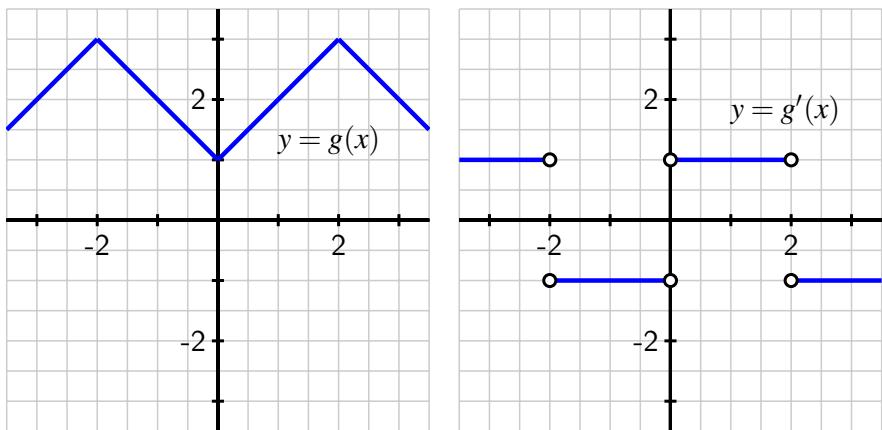
Answer.

- a. g is linear.
- b. On $-3.5 < x < -2$, $-2 < x < 0$ and $2 < x < 3.5$.
- c. At $x = -2, 0, 2$; g must have sharp corners at these points.
- d.



Solution.

- a. Since $g'(x)$ is constant (with value 1) on the interval $0 < x < 2$, it follows that g is linear on that same interval, since g is increasing at a constant rate.
- b. On $-3.5 < x < -2$, we also expect g to be linear with slope 1, while on $-2 < x < 0$ and $2 < x < 3.5$, g will be linear with slope -1; in each case this is true because the value of $g'(x)$ is constant with the noted value on the interval.
- c. From the given graph of $g'(x)$, we observe that g' is undefined at $x = -2, 0, 2$. Since we have been given that g is a continuous function, we can conclude that g must have sharp corners on its graph at these points. Moreover, that makes sense in light of our earlier observations that show g has constant slope on the intervals that connect at $x = -2, 0, 2$ and the graph jumps from having slope 1 to -1, and then back, and so on.
- d.



10. For each graph that provides an original function $y = f(x)$ in Figure 1.5.7, your task is to sketch an approximate graph of its derivative function, $y = f'(x)$, on the axes immediately below. 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' , you should label that accordingly.

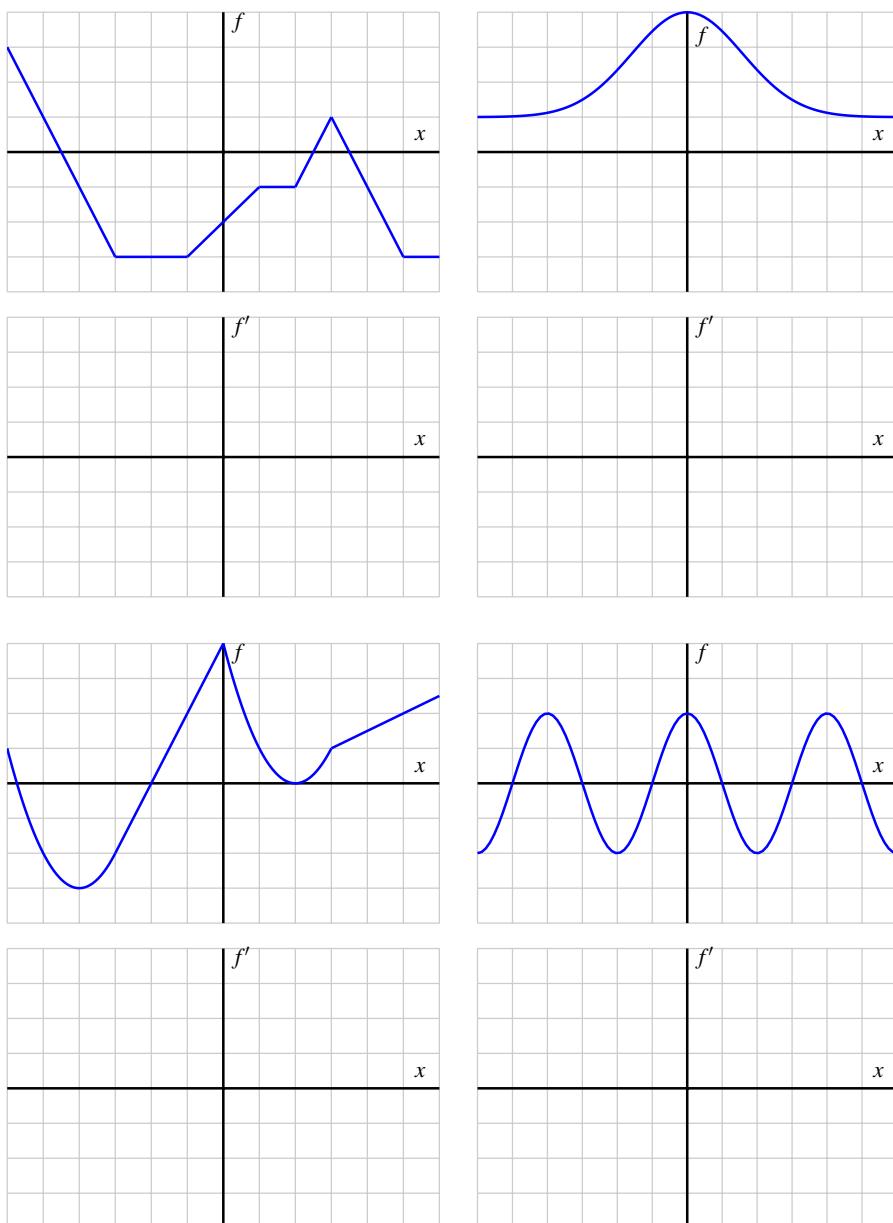
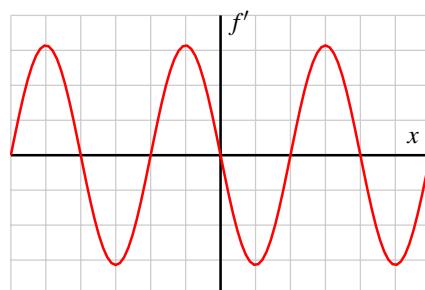
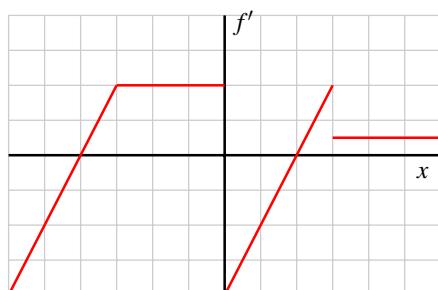
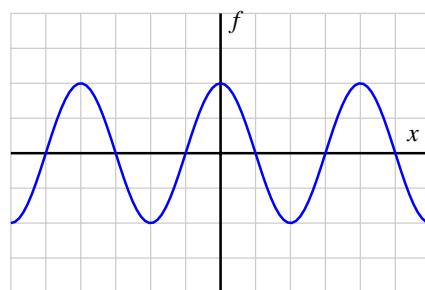
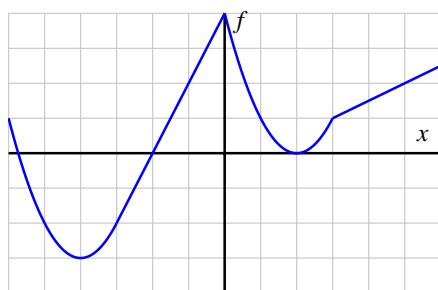
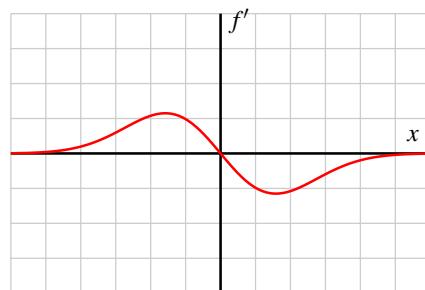
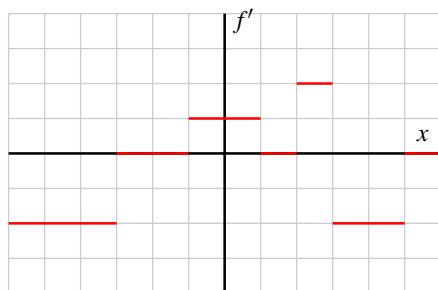
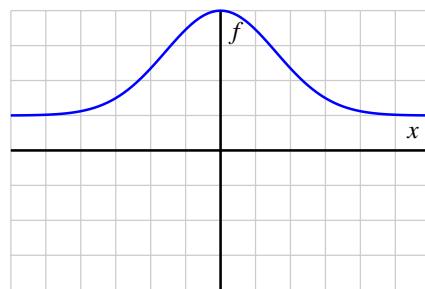
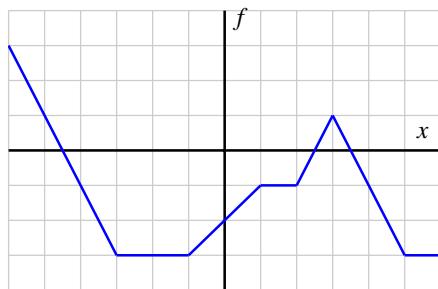
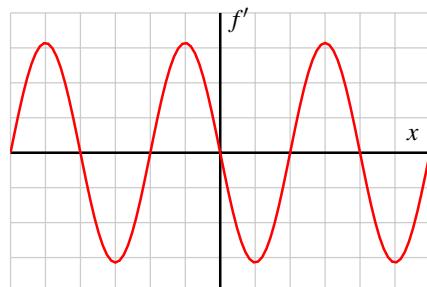
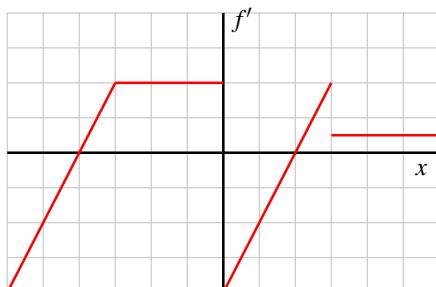
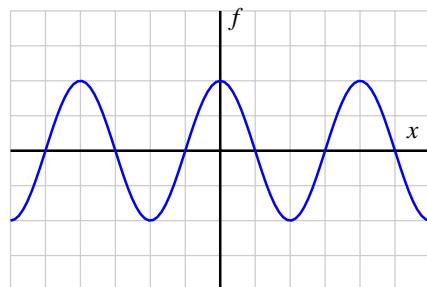
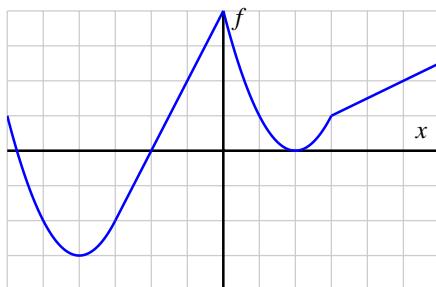
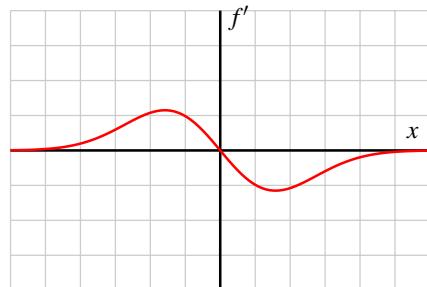
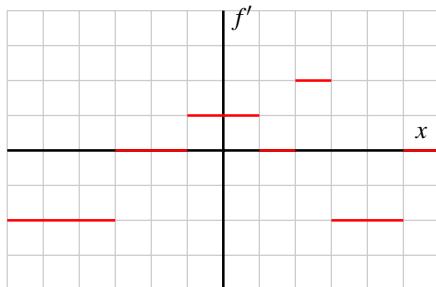
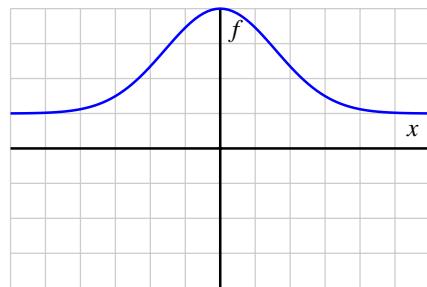
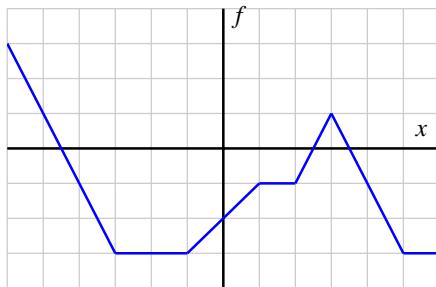


Figure 1.5.7: Graphs of $y = f(x)$ and grids for plotting the corresponding graph of $y = f'(x)$.

Answer.



Solution. At any point where there is a jump in the graph of the derivative, the derivative is undefined. Normally we would draw an open circle at each end of the graph, but those are omitted here for convenience of plotting.



1.6 Interpreting, estimating, and using the derivative

Motivating Questions

- 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?
-

A powerful feature of mathematics is that it can be studied both as an abstract discipline and as an applied one. For instance, calculus can be developed almost entirely as an abstract collection of ideas that focus on properties of functions. At the same time, if we consider functions that represent meaningful processes, calculus can describe our experience of physical reality. We have already seen that for the position function $y = s(t)$ of a ball being tossed straight up in the air, the derivative of the position function, $v(t) = s'(t)$, gives the ball's velocity at time t .

In this section, we investigate several functions with specific physical meaning, and consider 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.

Example 1.6.1. 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. Suppose 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 1.6.2 below. Three important points are labeled on the graph; where the curve looks linear, you may assume that it is indeed a straight line.

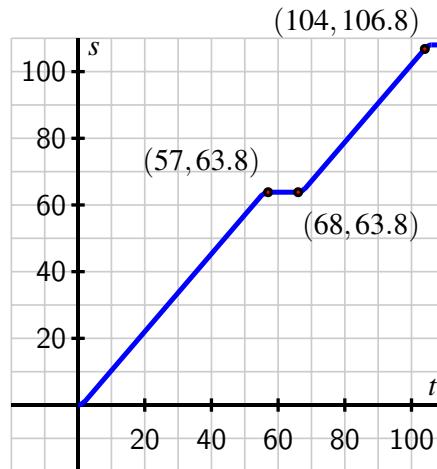


Figure 1.6.2: The graph of $y = s(t)$, the position of the car along highway 46, which tells its distance in miles from Gackle, ND at time t in minutes.

- 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 $[68, 104]$.
- 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?
- Find the average rate of change of the car's position on the interval $[68, 104]$. Include units in your answer.
- 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.

Hint.

- When is the car moving? When is the car stopped? How do you know? (What features of the graph tell you this information?)
- Units of slope should be “units of s per unit of t ”. Think about what the slope of a secant line represents.
- Use the formula for average rate of change.
- Close to the time $t = 80$ minutes, the function is linear. What does this mean for the instantaneous rate of change of the car's position at this instant?

Answer.

- The car drives away from Gackle at a constant speed for 57 minutes, stops for 11 minutes (on the interval $[57, 68]$), and then resumes driving away from Gackle at a constant speed for another 36 minutes (on the interval $[68, 104]$).
- The slope between these points is $m = \frac{43}{57} \approx 0.75$ miles per minute.
- $AV_{[68,104]} = \frac{43}{36} \approx 1.194$ miles per minute.
- $s'(80) = \frac{43}{36}$ miles per minute, the speed of the car after 80 minutes of driving.

Solution.

- For the first 57 minutes, the car travels away from Gackle at a constant speed. Since it travels a total of 63.8 miles during this time, its velocity is $\frac{63.8}{57}$ miles per minute, or approximately 67.16 miles per hour. The car is stopped between minutes 57 and 68, and then resumes driving away from Gackle, again at a constant speed. Between 68 and 104 minutes, the car travels $106.8 - 63.8 = 43$ miles, so its (average) velocity during this time is $\frac{43}{104-68} = \frac{43}{36}$ miles per minute, or about 71.67 miles per hour.
- The slope of the line between the points $(57, 63.8)$ and $(104, 106.8)$ is $m = \frac{106.8-63.8}{104-57} = \frac{43}{57} \approx 0.754$ miles per minute. This value represents the average speed of the car between 57 and 104 minutes of driving. Because the car is stopped for the first 11 minutes of this time interval, the average speed on this interval is down to approximately 45 miles per hour.
- The average rate of change of the car's position on the interval $[68, 104]$ is

$$AV_{[68,104]} = \frac{106.8 - 63.8}{104 - 68} = \frac{43}{36} \approx 1.194$$

miles per minute. On this interval, the car travels at an average speed of 72 miles per hour.

- At $t = 80$ minutes, we are solidly in the middle of the time interval $[68, 104]$, over which the car appears to be traveling at a constant velocity of about 72 miles per hour. A constant velocity (and in general, a linear position function) on an interval means the instantaneous velocity of the car at a point on that interval is the same as the average velocity of the car over that interval. Therefore, the instantaneous rate of change of the car's position at the moment $t = 80$ minutes is about 72 miles per hour, or 1.2 miles per minute.

1.6.1 Units of the derivative function

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

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

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 $(a, f(a))$.

We also sometimes write $\frac{df}{dx}$ or $\frac{dy}{dx}$ instead of $f'(x)$, and these alternate notations help us see the units (and thus the meaning) of the derivative as *the instantaneous rate of change of f with respect to x* . The units on the slope of the secant line, $\frac{f(x+h)-f(x)}{h}$, are “units of y per unit of x ,” and when we take the limit as h goes to zero, the derivative $f'(x)$ has the same units: units of y per unit of x . It is helpful to remember that the units on the derivative function are “units of output per unit of input,” for the variables of the original function.

For example, suppose that the function $y = P(t)$ measures the population of a city (in thousands) at the start of year t (where $t = 0$ corresponds to 2010 AD). 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.

1.6.2 Toward more accurate derivative estimates

Recall that to estimate the value of $f'(x)$ at a given x , we calculate a *difference quotient* $\frac{f(x+h)-f(x)}{h}$ with a relatively small value of h . We should use both positive and negative values of h in order to account for the behavior of the function on both sides of the point of interest. To that end, we introduce the notion of a *central difference* and its role in estimating derivatives.

Example 1.6.3. 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 $f'(2)$.

Hint. Try using a difference quotient with $h = \pm 1$.

Answer. $f'(2) \approx 0.5625$

Solution. We know that $f'(2) = \lim_{h \rightarrow 0} \frac{f(2+h)-f(2)}{h}$. But since we don’t have a graph or a formula for the function, we can neither sketch a tangent line nor evaluate the limit algebraically. 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

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

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

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

The intuitive approach to average the two estimates found in Example 1.6.3 is in fact the best possible way to estimate a derivative when we have just two function values for f on opposite sides of the point of interest.

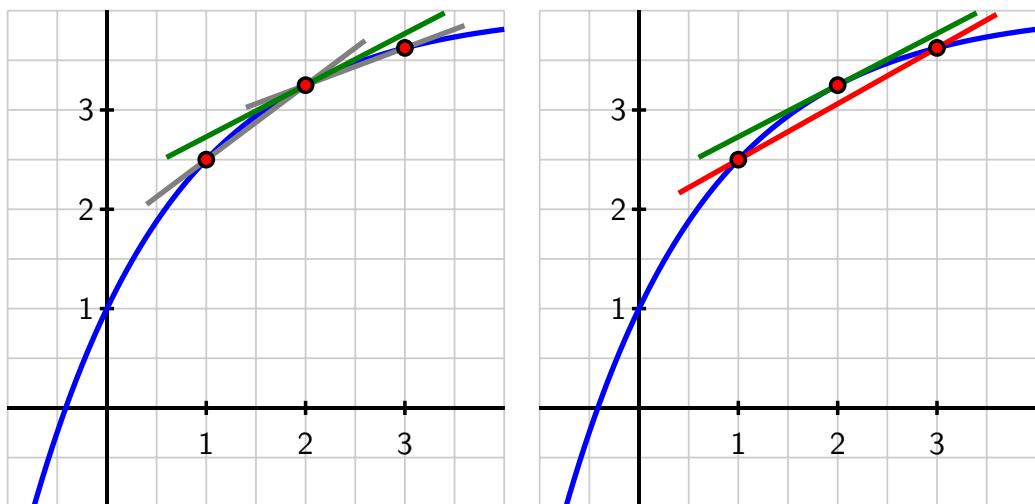


Figure 1.6.4: At left, the graph of $y = f(x)$ along with the secant line through $(1, 2.5)$ and $(2, 3.25)$, the secant line through $(2, 3.25)$ and $(3, 3.625)$, as well as the tangent line through $(2, 3.25)$. At right, the same graph of $y = f(x)$ along with the secant line through $(1, 2.5)$ and $(3, 3.625)$, plus the tangent line through $(2, 3.25)$.

To see why, we think about the diagram in Figure 1.6.4 above. 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 slope overestimates the slope of the tangent line at $(2, f(2))$, while the second slope underestimates $f'(2)$. On the right, 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 same value as the average of the forward and backward differences (and it is straightforward to explain why this always holds). The central difference yields a very good approximation to the derivative's value, because it yields a line closer to being parallel to the tangent line.

Central Difference Approximation

The **central difference approximation** to the value of the first derivative is given by

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

This quantity measures the slope of the secant line to $y = f(x)$ through the points

$(a - h, f(a - h))$ and $(a + h, f(a + h))$.

Example 1.6.5. 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 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

Table 1.6.6: Temperature data in degrees Fahrenheit.

- Use a central difference to estimate the instantaneous rate of change of the temperature of the potato at $t = 30$. Include units in your answer.
- Use a central difference to estimate the instantaneous rate of change of the temperature of the potato at $t = 60$. Include units in your answer.
- Without doing any calculation, which do you expect to be greater: $F'(75)$ or $F'(90)$? Why?
- 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?
- 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.

Hint.

- Think about quantities such as $\frac{F(45) - F(30)}{45 - 30}$.
- See the note in (a).
- Is F changing faster at $t = 75$ or at $t = 90$?
- Remember that the units on F' will be “degrees Fahrenheit per minute.”
- Be careful to distinguish between the temperature, F , and the rate of change of temperature, F' , in your commentary.

Answer.

- $F'(30) \approx 3.85$ degrees per minute.
- $F'(60) \approx 1.56$ degrees per minute.

- c. $F'(75) > F'(90)$.
- d. The value $F(64) = 330.28$ is the temperature of the potato in degrees Fahrenheit at time 64, while $F'(64) = 1.341$ measures the instantaneous rate of change of the potato's temperature with respect to time at the instant $t = 64$, and its units are degrees per minute. Because at time $t = 64$ the potato's temperature is increasing at a rate of 1.341 degrees per minute, we expect that at $t = 65$, the temperature will be about 1.341 degrees greater than at $t = 64$, or in other words $F(65) \approx 330.28 + 1.341 = 331.621$ degrees. Similarly, at $t = 66$, two minutes have elapsed from $t = 64$, so we expect an increase of $2 \cdot 1.341$ degrees: $F(66) \approx 330.28 + 2 \cdot 1.341 = 332.962$ degrees.
- e. Throughout the time interval $[0, 90]$, the temperature F of the potato is increasing. But as time goes on, the rate at which the temperature is rising appears to be decreasing. That is, while the values of F continue to get larger as time progresses, the values of F' are getting smaller (while still remaining positive). We thus might say that "the temperature of the potato is increasing, but at a decreasing rate."

Solution.

- a. Using the central difference, we find that

$$F'(30) \approx \frac{F(45) - F(15)}{45 - 15} = \frac{296 - 180.5}{30} = 3.85$$

degrees per minute.

- b. Using the central difference, we find that

$$F'(60) \approx \frac{F(75) - F(45)}{45 - 15} = \frac{342.8 - 296}{30} = 1.56$$

degrees per minute.

- c. Over each subsequent time interval, we see that the amount of increase in the potato's temperature gets less and less, thus we expect the value of $F'(t)$ to get smaller and smaller as time goes on. We therefore expect $F'(75) > F'(90)$.
- d. The value $F(64) = 330.28$ is the temperature of the potato in degrees Fahrenheit at time 64, while $F'(64) = 1.341$ measures the instantaneous rate of change of the potato's temperature with respect to time at the instant $t = 64$, and its units are degrees per minute. Because at time $t = 64$ the potato's temperature is increasing at a rate of 1.341 degrees per minute, we expect that at $t = 65$, the temperature will be about 1.341 degrees greater than at $t = 64$, or in other words $F(65) \approx 330.28 + 1.341 = 331.621$ degrees. Similarly, at $t = 66$, two minutes have elapsed from $t = 64$, so we expect an increase of $2 \cdot 1.341$ degrees: $F(66) \approx 330.28 + 2 \cdot 1.341 = 332.962$ degrees.
- e. Throughout the time interval $[0, 90]$, the temperature F of the potato is increasing. But as time goes on, the rate at which the temperature is rising appears to be decreasing. That is, while the values of F continue to get larger as time progresses, the values of F' are getting smaller (while still remaining positive). We thus might say that "the temperature of the potato is increasing, but at a decreasing rate."

Example 1.6.7. A company manufactures rope, and the total cost of producing r feet of rope is $C(r)$ dollars.

- a. What does it mean to say that $C(2000) = 800$?
- b. What are the units of $C'(r)$?
- c. 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.
- d. Do you think $C'(2000)$ is less than, equal to, or greater than $C'(3000)$? Why?
- e. 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?

Hint.

- a. The cost of producing 2000 feet of rope is . . .
- b. Remember that the units on any derivative are “units of output per unit of input.”
- c. How much do you expect C to increase for each additional foot away from 2000? By how many feet will the input increase to get to 2100?
- d. In manufacturing processes, there is often a decrease in cost per unit as the number of units increases.
- e. Is it possible for the total cost function to be decreasing at some point?

Answer.

- a. It costs \$800 to make 2000 feet of rope.
- b. “dollars per foot.”
- c. $C(2100) \approx 835$.
- d. Either $C'(2000) = C'(3000)$ or $C'(2000) > C'(3000)$.
- e. Impossible. The total cost function $C(r)$ can never decrease.

Solution.

- a. $C(2000) = 800$ means that it costs \$800 to make 2000 feet of rope.
- b. The units of $C'(r)$ are “dollars per foot.”

- c. 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.

- d. 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.
- e. 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.

Example 1.6.8. 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 in 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).

Hint.

- Try a central difference.
- What is happening when the car is traveling at 80 km/hr?
- Remember that units on the derivative are "units of output per unit of input."

Answer.

- $f'(90) \approx 0.0006$ liters per kilometer per kilometer per hour.
- At 80 kilometers per hour, the car is using fuel at a rate of 0.015 liters per kilometer.
- When the car is traveling at 90 kilometers per hour, its rate of fuel consumption per kilometer is increasing at a rate of 0.0006 liters per kilometer per kilometer per hour.

Solution.

- a. Using a central difference, we have

$$f'(90) \approx \frac{f(100) - f(80)}{100 - 80} = \frac{0.027 - 0.015}{20} = \frac{0.012}{20} = 0.0006$$

which tells us that $f'(90) \approx 0.0006$ liters per kilometer per kilometer per hour.

- b. When the car is traveling at 80 kilometers per hour, it is using fuel at a rate of 0.015 liters per kilometer. That is, at the given speed, for each additional kilometer the car travels, it uses an additional 0.015 liters of fuel.
- c. To say that $f'(90) = 0.0006$ liters per kilometer per kilometer per hour means that when the car is traveling at 90 kilometers per hour, its rate of fuel consumption per kilometer is increasing at a rate of 0.0006 liters per kilometer per kilometer per hour. If we increase our speed from 90 to 91 km/hr, we would expect our rate of fuel consumption to rise by 0.0006 liters for each additional kilometer driven.

In Section 1.5, we learned how to use the graph of a given function f to plot the graph of its derivative, f' . It is important to remember that when we do so, the scale and the units on the vertical axis often have to change to represent f' . For example, suppose that $P(t) = 400 - 330e^{-0.03t}$ tells us the temperature in degrees Fahrenheit of a potato in an oven at time t in minutes. In Figure 1.6.9 below, we sketch the graph of P on the left and the graph of P' on the right.

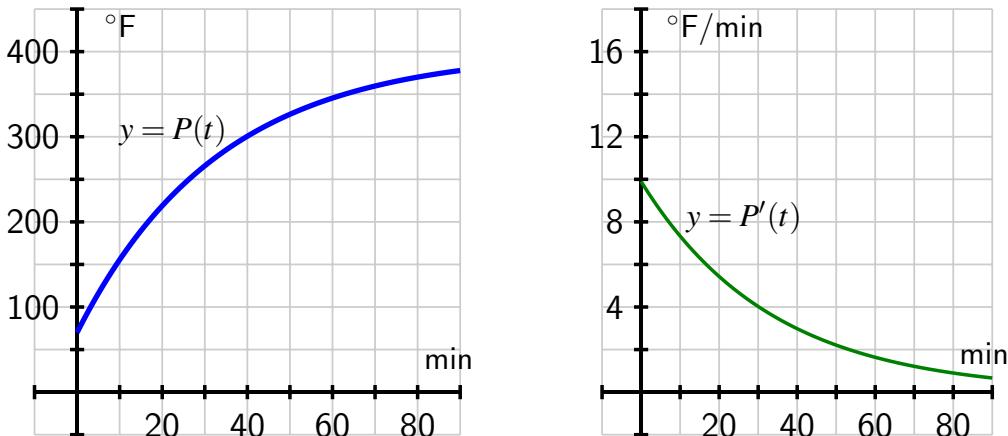


Figure 1.6.9: Plot of $P(t) = 400 - 330e^{-0.03t}$ at left, and its derivative $P'(t)$ at right.

Notice that both the size and the units of the vertical scales differ, as the units of P are °F, while those of P' are °F/min.

1.6.3 Summary

- The derivative of a given function $y = f(x)$ measures the instantaneous rate of change of the output variable with respect to the input variable.
- The units on the derivative function $y = f'(x)$ are units of $f(x)$ per unit of x . Again, this measures how fast the output of the function f changes when the input of the function changes.
- The central difference approximation to the value of the first derivative is given by

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

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.

1.6.4 Exercises

1. (A cooling cup of coffee) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (A cost function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Weight as a function of calories) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Displacement and velocity) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. 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?

Answer.

- a. $F'(10) \approx -3.33592$.
- b. The coffee's temperature is decreasing at about 3.33592 degrees per minute.
- c. $F'(20)$.
- d. We expect F' to get closer and closer to 0 as time goes on.

Solution.

- a. Using the central difference formula, we find that

$$F'(10) \approx \frac{F(10.01) - F(9.99)}{0.02} = \frac{(75 + 110e^{-0.05 \cdot 10.01}) - (75 + 110e^{-0.05 \cdot 9.99})}{0.02} \approx -3.33592.$$

- b. Since F is measured in degrees Fahrenheit and t in minutes, we have $F'(10) \approx -3.33592$ degrees F per minute. This means that at the instant $t = 10$, the temperature of the coffee is decreasing at an instantaneous rate of about 3.33592 degrees per minute. Said differently, in the next minute, we'd expect the temperature of the coffee to drop about 3.33592 degrees.
- c. We expect the coffee's temperature to decrease quickly early on when the difference between its temperature and the surroundings is greatest, and as time goes on for the coffee's temperature to approach that of the surrounding room. Keeping in mind that F is an always decreasing function so that F' is always negative, we observe that $F'(10)$ should be less (more negative) than $F'(20)$. These observations can also be confirmed by plotting a graph of F .
- d. We have noted that F' is always negative, and that we expect F' to get closer and closer to 0 as time goes on (that is, F' is increasing). As such, we expect F' to look something like the graph of $y = -e^{-x}$. This is reasonable since $-e^{-x}$ is also an exponential function, is always negative, is always increasing, and approaches 0 as x increases.
6. 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 $T = f(q)$.
- What does it mean to say $f(50) = 0.75$? Write a complete sentence to explain, using correct units.
 - A person's sensitivity, s , to the drug is defined by the function $s(q) = f'(q)$. What are the units of sensitivity?
 - 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).

Answer.

- If a patient takes a dose of 50 ml of a drug, the patient will experience a body temperature change of 0.75 degrees F.
- "degrees Fahrenheit per milliliter."
- For a patient taking a 50 ml dose, adding one more ml to the dose leads us to expect a temperature change that is about 0.02 degrees less than the temperature change induced by a 50 ml dose.

Solution.

- To say that $f(50) = 0.75$ means that if a patient takes a dose of 50 ml of a drug, the patient will experience a body temperature change of 0.75 degrees F.
- Since the units of f are degrees F and the units of q are ml, it follows that the units of $s(q) = f'(q)$ are "degrees Fahrenheit per milliliter."
- If $f'(50) = -0.02$ degrees Fahrenheit per milliliter, this means that the instantaneous rate of change of the temperature change function is -0.02 degrees F per ml at the dose-level of 50 ml. This means that for a patient taking a 50 ml dose, adding one more ml to the dose leads us to expect a temperature change that is about 0.02 degrees less than the temperature change induced by a 50 ml dose.

7. 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 $t = 2$.

- When is the ball's velocity greatest?
- Determine the value of $v'(1)$. Justify your thinking.
- 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 $t = 1$?
- What is the physical meaning of the function $v'(t)$?

Answer.

- $t = 0$.
- $v'(1) = -32$.
- “feet per second per second”; $v'(1) = -32$ tells us that the ball's velocity is decreasing at a rate of 32 feet per second per second.
- The acceleration of the ball.

Solution.

- Observe that $v(t) = 16 - 32t$ is a linear function with negative slope -32 , and thus decreasing. Since the function is of interest from $t = 0$ to $t = 2$, the velocity is greatest at the left endpoint, when $t = 0$. This greatest velocity is $v(0) = 16$ feet per second.
 - Since v is a linear function, its slope is constant, and hence the slope of every tangent line to the function is the same as the slope of the line the function generates. Thus, $v'(1) = -32$.
 - Because v is measured in feet per second and t in seconds, the units on $v'(1)$ are “feet per second per second.” Knowing that $v'(1) = -32$, this tells us that the ball's velocity is decreasing at a rate of 32 feet per second per second; that is, for whatever the ball's velocity is at $t = 1$, we would expect the velocity to decrease by (exactly) 32 feet per second over the next one second. (Normally this total decrease would be approximate, not exact, but since the velocity function is linear, we can compute the exact total change.)
 - The instantaneous rate of change of velocity is precisely the acceleration of the ball. Indeed, $v'(t)$ tells us how the ball's velocity changes, and that it does so at a constant rate of -32 feet per second per second. This is also consistent with the fact that acceleration due to gravity is constant.
8. 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 $V = h(m)$.
- 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?
 - 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.
 - Which value do you expect to be greater: $h'(30000)$ or $h'(80000)$? Why?
 - 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.

Answer.

- $AV_{[40000, 55000]} \approx -0.153$ dollars per mile.
- $h'(55000) \approx -0.147$ dollars per mile. During 550001st mile, we expect the car's value to drop by 0.147 dollars.
- $h'(30000) < h'(80000)$.
- The graph of h might have the general shape of the graph of $y = e^{-x}$ for positive values of x : always positive, always decreasing, and bending upwards while tending to 0 as x increases.

Solution.

- To find the average rate of change of h on the interval $[40000, 55000]$, we divide the change in h values (in dollars) on this interval by the change in m values (in miles) to obtain

$$AV_{[40000, 55000]} = \frac{h(55000) - h(40000)}{55000 - 40000} = \frac{13200 - 15500}{15000} = -\frac{2300}{15000} \approx -0.153$$

dollars per mile. This tells us that from the mileage of 40000 to 50000 miles, on average the car's value decreases by 0.153 dollars per mile driven.

- Since 40000 and 70000 are equidistant from 55000, we use the central difference quotient to approximate the derivative of h at 55000 to obtain

$$h'(55000) \approx \frac{h(70000) - h(40000)}{70000 - 40000} = \frac{11100 - 15500}{30000} = -\frac{4400}{30000} \approx -0.147$$

dollars per mile. This approximation tells us that at the instant the car has been driven 55000 miles, the car's value is decreasing at an instantaneous rate of 0.147 dollars per mile. That is, during 550001st mile, we'd expect the car's value to drop by 0.147 dollars.

- As we pile up miles on our car, we expect the value to go down. However, we should expect the depreciation to be more extreme at first, but less in magnitude as we drive more miles and our car gets older. Note, though, the the value of the car is always decreasing and $h'(m)$ is always negative. Thus, we should expect that the rate of change of h is increasing (becoming less negative) as m increases and that $h'(30000) < h'(80000)$.
- The value of a car is never negative. So we know that $h(m) \geq 0$ for all m . A new car decreases in value very quickly at first, and as the car gets older with more mileage, its value will depreciate to near zero. So h is a decreasing function. Since the value of our car always decreases, it follows that $h'(m) < 0$ for all $m > 0$. The fact that the depreciation slows as m increases implies that h' is an increasing (getting less negative) function. Since h' is increasing, it follows that h is bends upward; we will study this issue more in Section 1.6. So the graph of h might have the general shape of the graph of $y = e^{-x}$ for positive values of x .

1.7 The second derivative

Motivating 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 (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 of the second derivative? How do they help us understand the rate of change of the rate of change?
-

Given a differentiable function $y = f(x)$, we know that its derivative, $y = f'(x)$, is a related function whose output at $x = a$ 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.

At a point where $f'(x)$ is positive, the slope of the tangent line to f is positive. Therefore, on an interval where $f'(x)$ is positive, the function f is increasing (or rising). Similarly, if $f'(x)$ is negative on an interval, the graph of f is decreasing (or falling).

The derivative of f tells us not only *whether* the function f is increasing or decreasing on an interval, but also *how* the function f is increasing or decreasing. Look at the two tangent lines shown in Figure 1.7.1. We see that at point A the value of $f'(x)$ is positive and relatively close to zero, and at that point the graph is rising slowly. By contrast, at point B , the derivative is negative and relatively large in absolute value, and f is decreasing rapidly at B .

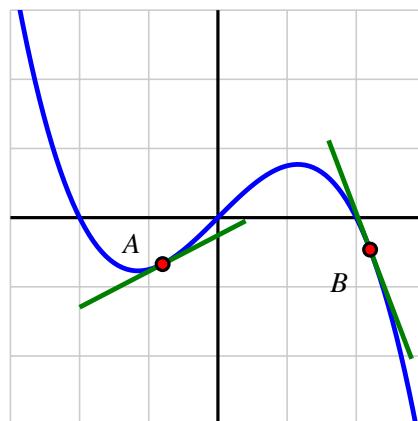


Figure 1.7.1: Two tangent lines on a graph.

Besides asking whether the value of the derivative function is positive or negative and whether it is large or small, we can also ask “how is the derivative changing?”

Because the derivative, $y = f'(x)$, is itself a function, we can consider taking its derivative — the derivative of the derivative — and ask “what does the derivative of the derivative tell us about how the original function behaves?” We start with an investigation of a moving object.

Example 1.7.2. 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 1.7.3. 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.

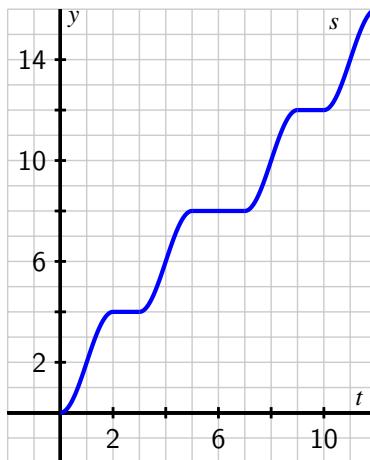


Figure 1.7.3: The graph of $y = s(t)$, the position of the car (measured in thousands of feet from its starting location) at time t in minutes.

- 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 $[0, 12]$.
- Sketch a careful, accurate graph of $y = s'(t)$.
- 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?
- 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 $y = v'(t)$. Write at least one sentence to explain how the behavior of $v'(t)$ is connected to the graph of $y = v(t)$.

Hint.

- Can you estimate the car's speed at different times? Does it ever stop or change direction? How can you tell?
- What are the units on s' ? When is the slope of the tangent line to s positive, zero, or negative? How big does it get?
- What are the units on s'' ? When is the value of the derivative positive, zero, or negative, and how does that relate to the direction the car is traveling at that time?
- When does your graph in (b) have positive slope? Negative slope? Zero slope?
- Look at your graph of $y = v(t)$ from (b). What do the slopes of the tangent lines to v tell you about the values of $v'(t)$?

Answer.

- The car starts out not moving and then speeds up for a minute as it travels about 1300 feet forward before starting to slow down in the second minute, coming to a stop at time $t = 2$ minutes at a point 4000 feet from its starting position. The car is stopped during the third minute. This three-minute pattern repeats for the full 12 minutes, at which point the car is 16,000 feet from its starting position, having always traveled in the same direction along the road.

b

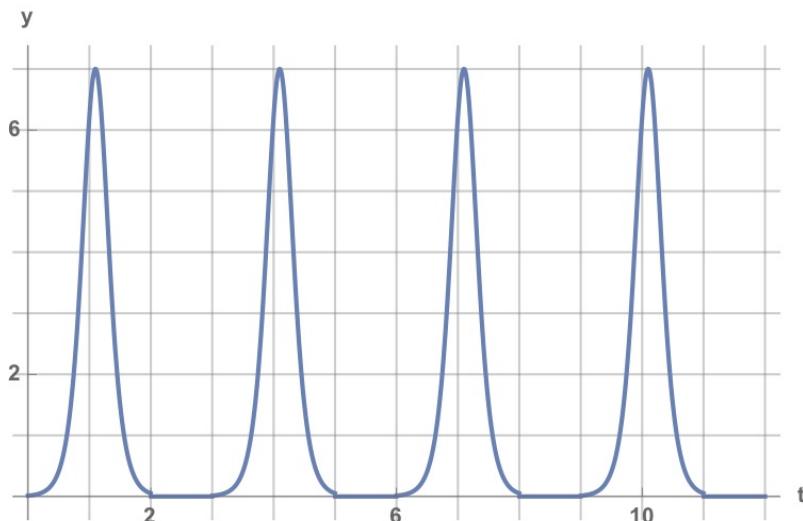


Figure 1.7.4: The graph of $y = s'(t)$, showing the velocity of the car, in thousands of feet per minute, after t minutes.

- c s' describes the velocity of the car, in 1000 ft/min, after t minutes of driving. The car moves forward when $s'(t)$ is positive, moves backward when $s'(t)$ is negative, and is stopped when $s'(t) = 0$.
- d v is increasing on the intervals $(0, 1.1)$, $(3, 4.1)$, $(6, 7.1)$, and $(9, 10.1)$. v is decreasing on the intervals $(1.1, 2)$, $(4.1, 5)$, $(7.1, 8)$, and $(10.1, 11)$. v is constant on the intervals $(2, 3)$, $(5, 6)$, $(8, 9)$, and $(11, 12)$.

e

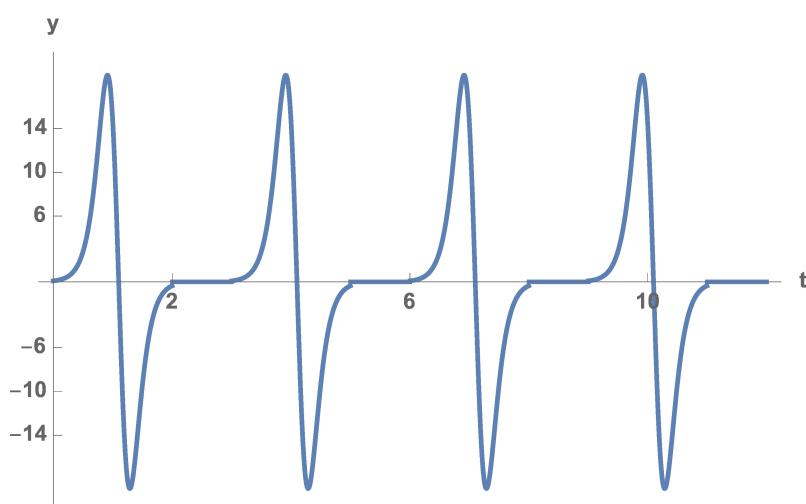


Figure 1.7.5: The graph of $y = v'(t)$, showing the acceleration of the car, in thousands of feet per minute per minute, after t minutes.

As seen in the graph above: v' is positive whenever v is increasing; v' is negative whenever v is decreasing; v' is zero whenever v is constant.

Solution.

- a At time $t = 0$, the car is at rest but gradually accelerates to a speed of about 6000 ft/min as it drives about 1300 feet during the first minute of travel. The car reaches its peak speed of about 7000 ft/min just after the time $t = 1$ minute (and again just after the points $t = 4$, $t = 7$, and $t = 10$ minutes). In the second minute, the car gradually slows back down, coming to a stop about 4000 feet from where it started. It waits there for a minute (between $t = 2$ minutes and $t = 3$ minutes) before continuing to drive in the same direction as before. At this point, the car again gradually accelerates to a speed of about 6000 ft/min by the end of the fourth minute, at which point it has driven around 5300 feet since starting out. During the fifth minute, the car gradually slows back to a stop after traveling an additional 2700 feet. This pattern of starts and stops continues for a total of 12 minutes, by which time the car has traveled a total of 16,000 feet from its starting point.

b

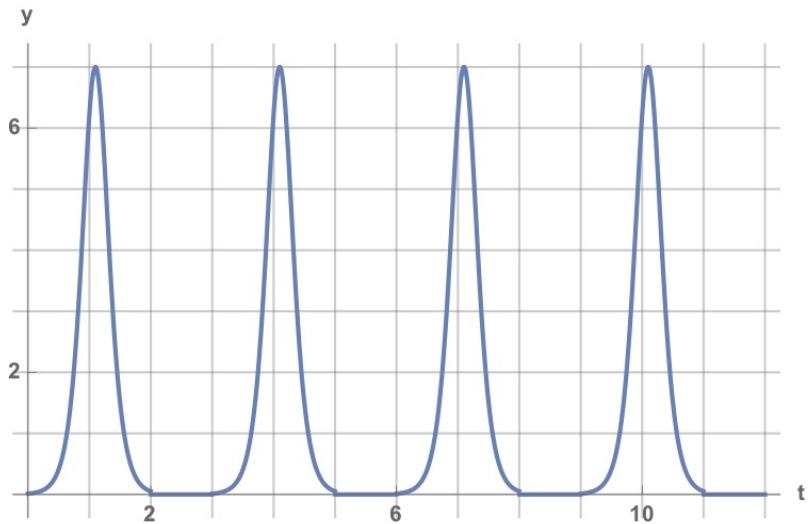


Figure 1.7.6: The graph of $y = s'(t)$, showing the velocity of the car, in thousands of feet per minute, after t minutes.

The car is stopped at $t = 0$ and $t = 12$ minutes, as well as on the intervals $(2, 3)$, $(5, 6)$, $(8, 9)$, and $(11, 12)$. The value of s' at these times is 0 ft/min. In the minute or so after each of the points $t = 0$, $t = 3$, $t = 6$, and $t = 9$, the car gradually accelerates to a speed of about 7000 ft/min, and then gradually slows back down, reaching a speed of 0 ft/min by the times $t = 2$, $t = 5$, $t = 8$, and $t = 11$ minutes. While the car is speeding up, the graph of $y = s'(t)$ has a positive slope; while the car is slowing down, the graph of $y = s'(t)$ has a negative slope. When the car is traveling at a constant speed (of 0 ft/min), the graph of $y = s'(t)$ is horizontal.

- c The derivative function s' describes the velocity of the car, in thousands of feet per minute, after t minutes of driving. The car moves forward when $s'(t)$ is positive, moves backward when $s'(t)$ is negative, and is stopped when $s'(t) = 0$.
- d v is increasing from 0 ft/min to 7000 ft/min approximately on the 66-second intervals $(0, 1.1)$, $(3, 4.1)$, $(6, 7.1)$, and $(9, 10.1)$. v is decreasing from 7000 ft/min to 0 ft/min approximately on the 54-second intervals $(1.1, 2)$, $(4.1, 5)$, $(7.1, 8)$, and $(10.1, 11)$. v is constant at 0 ft/min on the 1-minute intervals $(2, 3)$, $(5, 6)$, $(8, 9)$, and $(11, 12)$.

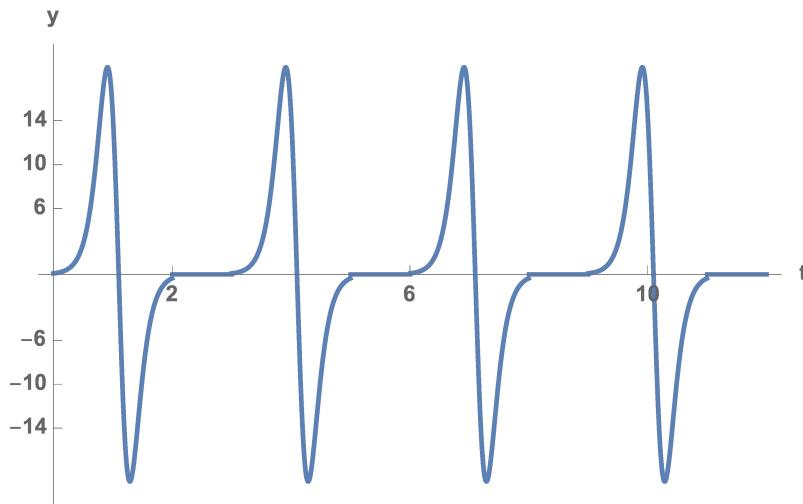


Figure 1.7.7: The graph of $y = v'(t)$, showing the acceleration of the car, in thousands of feet per minute per minute, after t minutes.

As seen in the graph above: v' is positive whenever v is increasing; v' is negative whenever v is decreasing; v' is zero whenever v is constant. Since the units of v' are thousands of feet per minute squared, we conjecture that the function v' models the acceleration of the car after t minutes of driving. Thus we have positive acceleration whenever the car is speeding up (increasing velocity), negative acceleration whenever the car is slowing down (decreasing velocity), and zero acceleration whenever the car travels at a constant speed (constant velocity).

1.7.1 Increasing, decreasing, or neither

So far, we have used the words *increasing* and *decreasing* intuitively to describe a function's graph. Here we define these terms more formally.

Increasing and Decreasing Functions

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. If the function has a derivative, the sign of the derivative tells us whether the function is increasing or decreasing.

Let f be a function that is differentiable on an interval (a, b) . It is possible to show that that f is increasing on (a, b) if and only if $f'(x) > 0$ for every x such that $a < x < b$; similarly, f is decreasing on (a, b) if and only if $f'(x) < 0$. If $f'(a) = 0$, then f is neither increasing nor

decreasing at $x = a$.

For example, the function pictured in Figure 1.7.8 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.

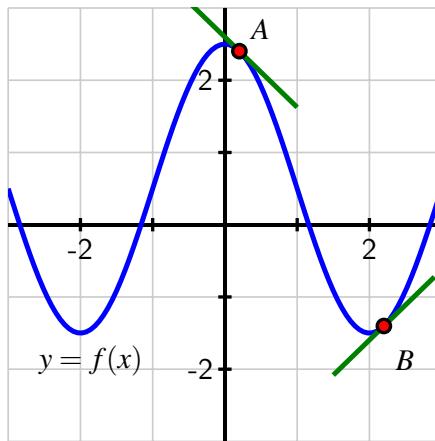


Figure 1.7.8: A function that is decreasing on the intervals $-3 < x < -2$ and $0 < x < 2$ and increasing on $-2 < x < 0$ and $2 < x < 3$.

1.7.2 The Second Derivative

We are now accustomed to investigating the behavior of a function by examining its derivative. The derivative of a function f is a new function given by the rule

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

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 f* , and denote the second derivative by $y = f''(x)$. Consequently, we will sometimes call f' “the first derivative” of f , rather than simply “the derivative” of f .

Second Derivative

The **second derivative** is defined by applying the limit definition of the derivative to the first derivative. That is,

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

We read “ $f''(x)$ ” as “ f -double prime of x ”, or as “the second derivative of f ”.

The meaning of the derivative function still holds, so when we compute $f''(x)$, 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. The second derivative will help us understand how the rate of change of the original function is itself changing.

1.7.3 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. There are three basic behaviors that an increasing function can demonstrate on an interval, as pictured below in Figure 1.7.9: the function can increase more and more rapidly, it can increase at the same rate, or it can 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.

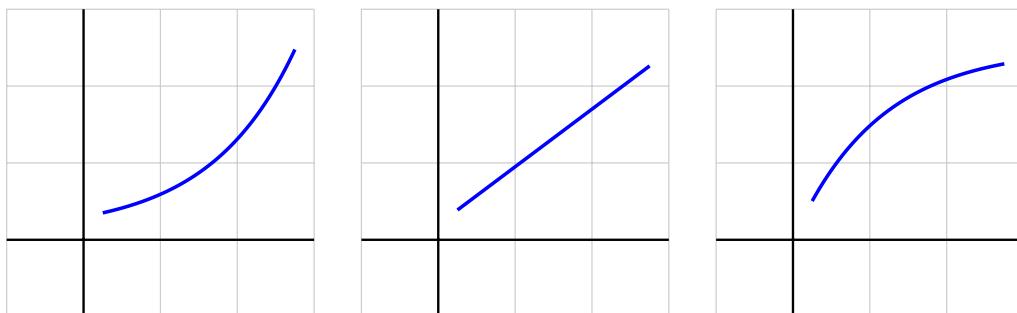


Figure 1.7.9: Three functions that are all increasing, but doing so at an increasing rate, at a constant rate, and at a decreasing rate, respectively.

On the leftmost curve in Figure 1.7.9, draw 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 1.7.9, observe that as x increases, the function increases, but the slopes of the tangent lines decrease. This function is *increasing at a decreasing rate*.

Similar options hold for how a function can decrease. Here we must be extra careful with our language, because decreasing functions involve negative slopes¹. When a function's values are negative, and those values get more negative as the input increases, the function must be decreasing.

¹Negative numbers present an interesting 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 “greater than” describes how numbers lie on a number line: $x > y$ provided that x lies to the right of y . So of course, -100 is less than -2 . Informally, it might be helpful to say that “ -100 is more negative than -2 .”

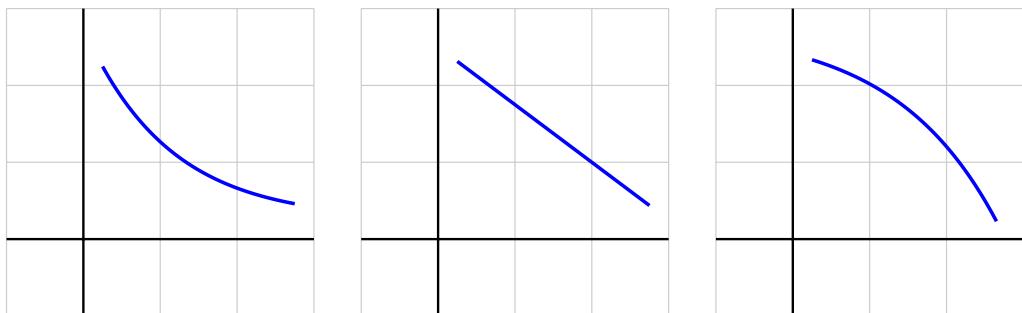


Figure 1.7.10: From left to right, three functions that are all decreasing, but doing so in different ways.

Now consider the three graphs shown above in Figure 1.7.10. The middle graph clearly depicts a function decreasing at a constant rate. Now, on the first curve, draw a sequence of tangent lines. We see that the slopes of these lines get closer to zero — meaning they 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*.

This leaves only the rightmost curve in Figure 1.7.10 to consider. For that function, the slopes of the tangent lines are negative throughout the pictured interval, but as we move from left to right, the slopes get more and more negative as they get steeper. Hence the slope of the curve is decreasing, and we say that the function is *decreasing at a decreasing rate*.

We now introduce the notion of *concavity* which provides simpler language to describe these behaviors. When a curve opens upward on a given interval, like the 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, like the parabola $y = -x^2$ or the negative exponential function $y = -e^x$, we say that the function is *concave down*. Concavity is linked to both the first and second derivatives of the function.

In Figure 1.7.11 below, we see two functions and a sequence of tangent lines to each. The lefthand plot shows a function that is concave up; observe that here the tangent lines always lie below the curve itself, and the slopes of the tangent lines are increasing as we move from left to right. In other words, the graph of $y = f(x)$ is concave up on the interval shown because its derivative, f' , is increasing on that interval. Similarly, the righthand plot in Figure 1.7.11 depicts a function that is concave down; in this case we see that the tangent lines always lie above the curve and the slopes of the tangent lines are decreasing as we move from left to right. The fact that its derivative, f' , is decreasing makes f concave down on the interval shown.

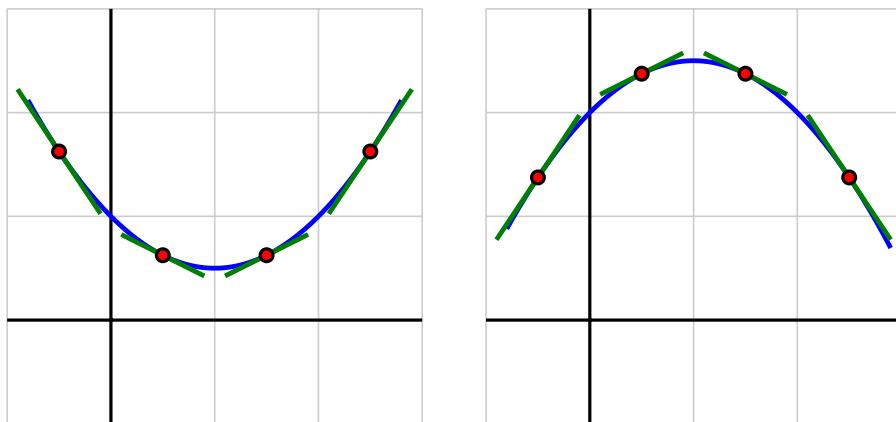


Figure 1.7.11: At left, a function that is concave up; at right, one that is concave down.

We state these most recent observations formally as the definitions of the terms *concave up* and *concave down*.

Concavity

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 1.7.12. 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 below in Figure 1.7.13. 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 $y = v'(t) = s''(t)$ earlier, in Example 1.7.2.

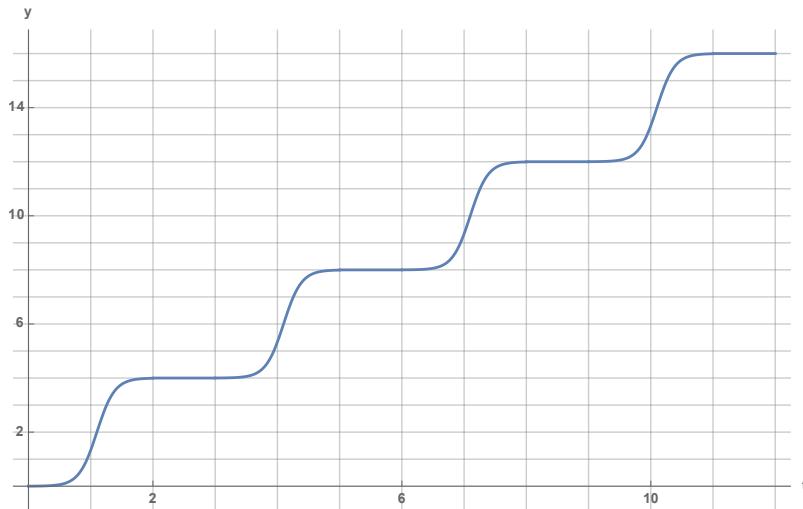


Figure 1.7.13: The graph of $y = s(t)$, the position of the car (measured in thousands of feet from its starting location) at time t in minutes.

- On what intervals is the position function $y = s(t)$ increasing? decreasing? Why?
- On which intervals is the velocity function $y = v(t) = s'(t)$ increasing? decreasing? neither? Why?
- 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.
- What can you say about s'' whenever s' is increasing? Why?
- 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 _____.

Hint.

- Remember that a function is increasing on an interval if and only if its first derivative is positive on the interval.
- See (a).
- Remember that the first derivative of a function measures its instantaneous rate of change.
- Think about how $s''(t) = [s'(t)]'$.
- Be very careful with your letters: s , v , and a .

Answer.

- Increasing: $0 < t < 2$, $3 < t < 5$, $6 < t < 8$, and $9 < t < 11$. Decreasing: never.
- Velocity is increasing on $0 < t < 1.1$, $3 < t < 4.1$, $6 < t < 7.1$, and $9 < t < 10.1$; $y = v(t)$ is decreasing on $1.1 < t < 2$, $4.1 < t < 5$, $7.1 < t < 8$, and $10.1 < t < 11$. Velocity is neither increasing nor decreasing (i.e. the velocity is constant) on $2 < t < 3$, $5 < t < 6$, $8 < t < 9$, and $11 < t < 12$.
- $a(t) = v'(t)$ and $a(t) = s''(t)$.
- $s''(t)$ is positive since $s'(t)$ is increasing.
- *increasing*.
 - *decreasing*.
 - *constant*.
 - *increasing*.
 - *decreasing*.
 - *constant*.
 - *concave up*.
 - *concave down*.
 - *linear*.

Solution.

- The position function $y = s(t)$ is increasing on the intervals $0 < t < 2$, $3 < t < 5$, $6 < t < 8$, and $9 < t < 11$ because $s'(t)$ is positive at every point in such intervals. Furthermore, $s(t)$ is never decreasing because its derivative is never negative.
- The velocity function $y = v(t)$ appears to be increasing on the intervals $0 < t < 1.1$, $3 < t < 4.1$, $6 < t < 7.1$, and $9 < t < 10.1$. This is because the curve $y = s(t)$ is concave up on these intervals, which corresponds to an increasing first derivative $y = s'(t)$. Similarly, $y = v(t)$ appears to be decreasing on the intervals $1.1 < t < 2$, $4.1 < t < 5$, $7.1 < t < 8$, and $10.1 < t < 11$. This is due to the curve $y = s(t)$ being concave down

on these intervals, corresponding to a decreasing first derivative $y = s'(t)$. On the intervals $2 < t < 3$, $5 < t < 6$, $8 < t < 9$, and $11 < t < 12$, the curve $y = s(t)$ is constant, and thus linear, so s is neither concave up nor concave down. On these intervals, then, the velocity function is constant.

- c. Since $a(t)$ is the instantaneous rate of change of $v(t)$, we can say $a(t) = v'(t)$. Moreover, because $v(t) = s'(t)$, it follows that $a(t) = v'(t) = [s'(t)]' = s''(t)$, so acceleration is the second derivative of position.
- d. Since $s''(t)$ is the first derivative of $s'(t)$, then whenever $s'(t)$ is increasing, $s''(t)$ must be positive.
- e. For the position function $s(t)$ with velocity $v(t)$ and acceleration $a(t)$,
 - on an interval where $v(t)$ is positive, $s(t)$ is *increasing*.
 - on an interval where $v(t)$ is negative, $s(t)$ is *decreasing*.
 - on an interval where $v(t)$ is zero, $s(t)$ is *constant*.
 - on an interval where $a(t)$ is positive, $v(t)$ is *increasing*.
 - on an interval where $a(t)$ is negative, $v(t)$ is *decreasing*.
 - on an interval where $a(t)$ is zero, $v(t)$ is *constant*.
 - on an interval where $a(t)$ is positive, $s(t)$ is *concave up*.
 - on an interval where $a(t)$ is negative, $s(t)$ is *concave down*.
 - on an interval where $a(t)$ is zero, $s(t)$ is *linear*.

Exploring the context of position, velocity, and acceleration is an excellent way to understand how a function, its first derivative, and its second derivative are related to one another. In Example 1.7.12 that we just finished, 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 the following are equivalent: on an interval where the graph of $y = f(x)$ is concave up, f' is increasing and f'' is positive. Likewise, on an interval where the graph of $y = f(x)$ is concave down, f' is decreasing and f'' is negative.

Example 1.7.14. 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 below in Table 1.7.15. Time t is measured in minutes. In Example 1.6.5 from Section 1.6, we computed approximations to $F'(30)$ and $F'(60)$ using central differences. Those values are provided in Table 1.7.16 below on the right, along with several others computed in the same way.

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

Table 1.7.15: Select values of $F(t)$.

t	$F'(t)$
0	NA
15	6.03
30	3.85
45	2.45
60	1.56
75	1.00
90	NA

Table 1.7.16: Select values of $F'(t)$.

- What are the units on the values of $F'(t)$?
- Use a central difference to estimate the value of $F''(30)$.
- In terms of the potato's temperature, what is the meaning of the value of $F''(30)$ that you have computed in (b)? 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.
- Overall, is the potato's temperature increasing at an increasing rate, increasing at a constant rate, or increasing at a decreasing rate? Why?

Hint.

- Remember that the derivative's units are "units of output per unit of input."
- Central difference: to estimate $g'(a)$, we can use

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

for an appropriate choice of h .

- For each of the values $F'(30)$ and $F''(30)$, think about what they tell you about expected upcoming behavior in $F(t)$ and $F'(t)$, respectively.
- Think about concavity.

Answer.

- Degrees Fahrenheit per minute.
- $F''(30) \approx -0.119$.
- At the moment $t = 30$, the temperature of the potato is 251 degrees; its temperature is rising at a rate of 3.85 degrees per minute; and the rate at which the temperature is rising is *falling* at a rate of 0.119 degrees per minute per minute.
- Increasing at a decreasing rate.

Solution.

- $F'(t)$ has units measured in degrees Fahrenheit per minute.
- Using a central difference,

$$F''(30) \approx \frac{F'(45) - F'(15)}{30} = \frac{2.45 - 6.03}{30} \approx -0.119.$$

- c. The values $F(30) = 251$, $F'(30) = 3.85$, and $F''(30) \approx -0.119$ (which is measured in degrees per minute per minute), tell us that at the moment $t = 30$ minutes: the temperature of the potato is 251 degrees, its temperature is rising at a rate of 3.85 degrees per minute, and the rate at which the temperature is rising is *falling* at a rate of 0.119 degrees per minute per minute. That is, even though the temperature is still rising, it is doing so at a slower and slower rate. At $t = 31$, we expect that the rate of increase of the potato's temperature would have dropped to about 3.73 degrees per minute.
- d. The potato's temperature is increasing at a decreasing rate because the values of the first derivative of F are falling. Equivalently, the value of $F''(t)$ is negative throughout the given time interval.

Example 1.7.17. This example builds on our experience and understanding of how to sketch the graph of $y = f'(x)$ given the graph of $y = f(x)$.

For each of the two functions graphed below in Figure 1.7.18, sketch the corresponding graphs of the first and second derivatives. In addition, for each, write several careful sentences in the spirit of those in Example 1.7.12 that connect the behaviors of f , f' , and f'' (or of g , g' , and g'' in the case of the second function). 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.”

but with the blanks filled in. The scale of the grids on the given graphs is 1×1 ; be sure to label the scale on each of the graphs you draw, even if it does not change from the original.

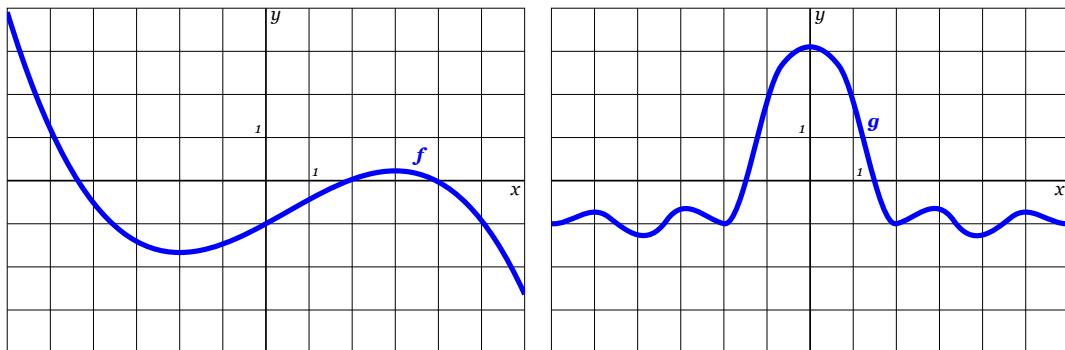
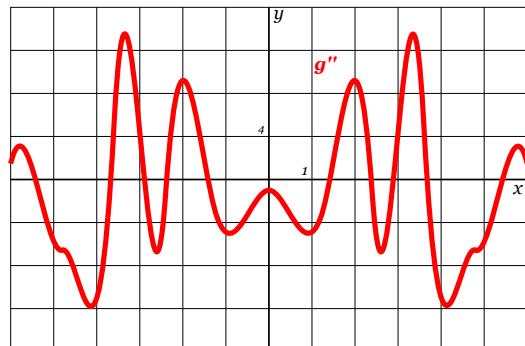
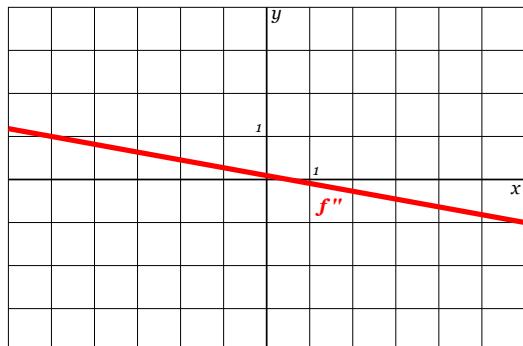
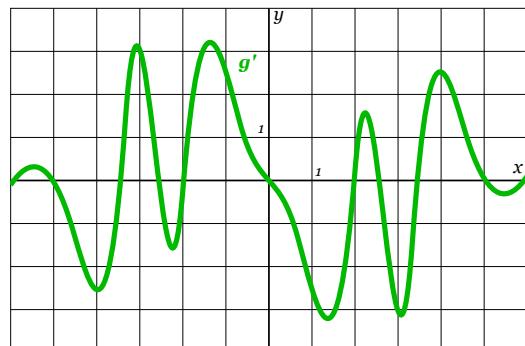
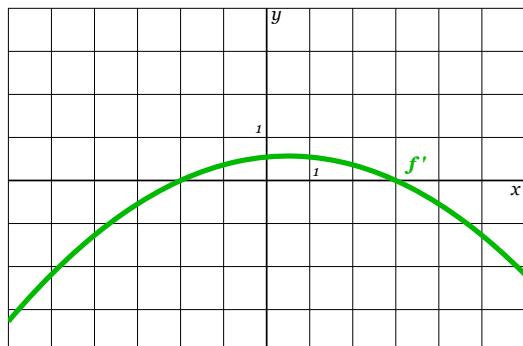
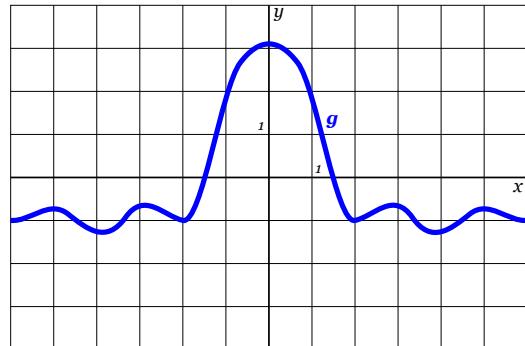
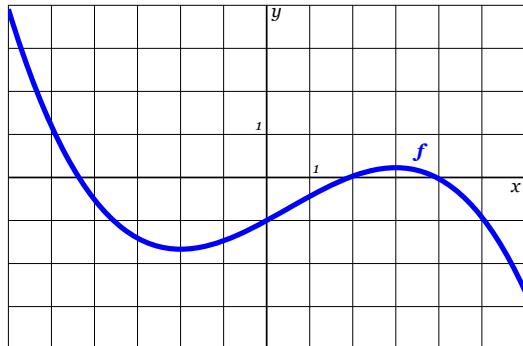


Figure 1.7.18: The graphs of $y = f(x)$ and $y = g(x)$, for two given functions f and g .

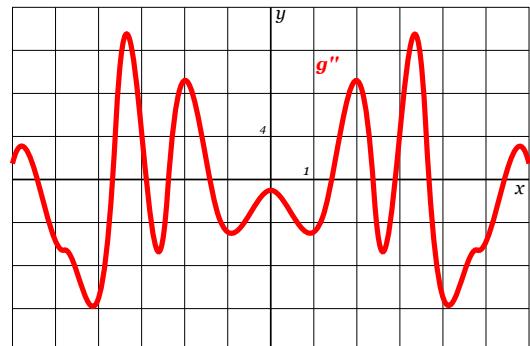
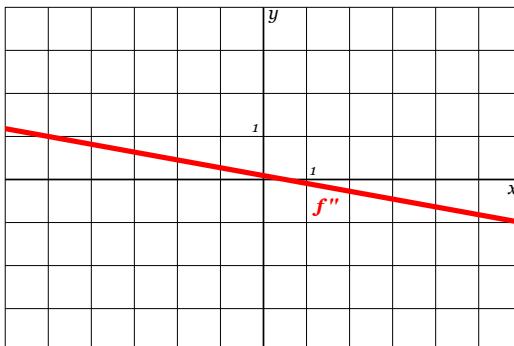
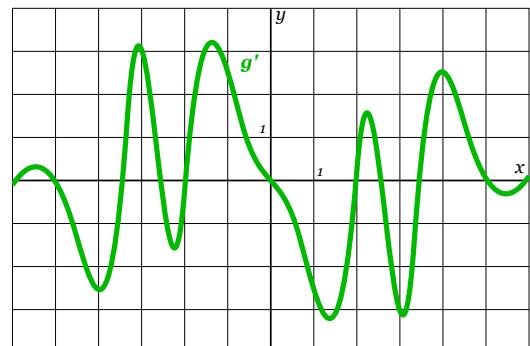
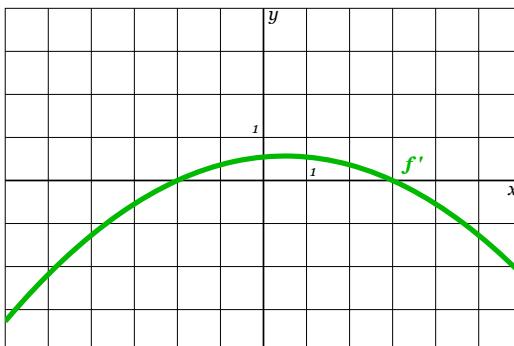
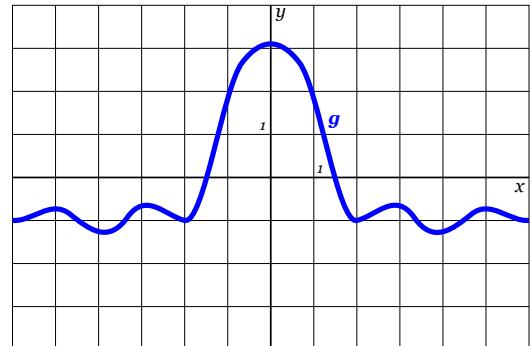
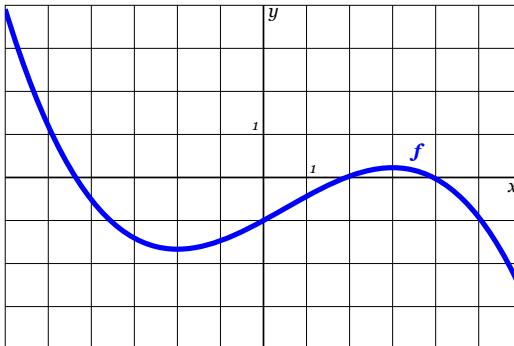
Hint. Remember that to plot $y = f'(x)$, it is helpful to first identify where $f'(x) = 0$.

Answer.



Notice the vertical scale on the graph of $y = g''(x)$ has changed, with each grid square now having height 4.

Solution. The graphs of $y = f'(x)$ and $y = f''(x)$ are plotted below the graph of $y = f(x)$ on the left. The graphs of $y = g'(x)$ and $y = g''(x)$ are plotted below the graph of $y = g(x)$ on the right. Notice the vertical scale on the graph of $y = g''(x)$ has changed, with each grid square now having height 4.



- The graph of $y = f(x)$ is decreasing and concave up on the interval $(-6, -2)$, which is connected to the fact that f'' is positive, and that f' is negative and increasing on the same interval.
- The graph of $y = f(x)$ is increasing and concave up on the interval $(-2, 0.5)$, which is connected to the fact that f'' is positive, and that f' is positive and increasing on the same interval.
- The graph of $y = f(x)$ is increasing and concave down on the interval $(0.5, 3)$, which is connected to the fact that f'' is negative, and that f' is positive and decreasing on the same interval.

- The graph of $y = f(x)$ is decreasing and concave down on the interval $(3, 6)$, which is connected to the fact that f'' is negative, and that f' is negative and decreasing on the same interval.
- The graph of $y = g(x)$ is increasing and concave up on the (approximate) intervals $(-6, -5.5)$, $(-3.5, -3)$, $(-2, -1.5)$, $(2, 2.2)$, and $(3.5, 4)$. This is connected to the fact that g'' is positive, and that g' is positive and increasing on the same intervals.
- The graph of $y = g(x)$ is increasing and concave down on the (approximate) intervals $(-5.5, -5)$, $(-3, -2.5)$, $(-1.5, 0)$, $(2.2, 2.5)$, and $(4, 5)$. This is connected to the fact that g'' is negative, and that g' is positive and decreasing on the same intervals.
- The graph of $y = g(x)$ is decreasing and concave down on the (approximate) intervals $(-5, -4)$, $(-2.5, -2.2)$, $(0, 1.5)$, $(2.5, 3)$, and $(5, 5.5)$. This is connected to the fact that g'' is negative, and that g' is negative and decreasing on the same intervals.
- The graph of $y = g(x)$ is decreasing and concave up on the (approximate) intervals $(-4, -3.5)$, $(-2.2, -2)$, $(1.5, 2)$, $(3, 3.5)$, and $(5.5, 6)$. This is connected to the fact that g'' is positive, and that g' is negative and increasing on the same intervals.

1.7.4 Summary

- 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. The sign of the second derivative tells us whether 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). Examples of functions that are everywhere concave up are $y = x^2$ and $y = e^x$; examples of functions that are everywhere concave down are $y = -x^2$ and $y = -e^x$.
- 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.

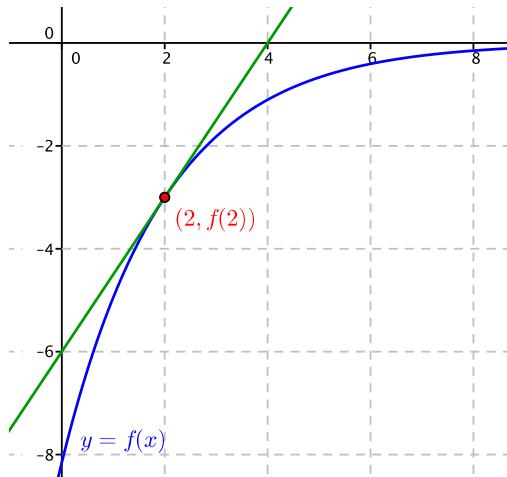
1.7.5 Exercises

1. (Comparing f, f', f'' values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Signs of f, f', f'' values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

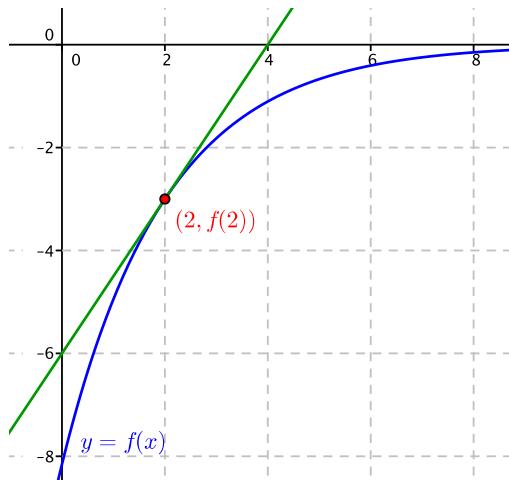
3. (Acceleration from velocity) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Rates of change of stock values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Interpreting a graph of f') This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. Suppose that $y = f(x)$ is a differentiable function for which the following information is known: $f(2) = -3$, $f'(2) = 1.5$, $f''(2) = -0.25$.
- Is f increasing or decreasing at $x = 2$? Is f concave up or concave down at $x = 2$?
 - Do you expect $f(2.1)$ to be greater than -3 , equal to -3 , or less than -3 ? Why?
 - Do you expect $f'(2.1)$ to be greater than 1.5 , equal to 1.5 , or less than 1.5 ? Why?
 - Sketch a graph of $y = f(x)$ near $(2, f(2))$ and include a graph of the tangent line.

Answer.

- f is increasing and concave down at $x = 2$.
- Greater.
- Less.
-

**Solution.**

- Since $f'(2)$ is positive, f is increasing at $x = 2$, and since $f''(2)$ is negative, f is concave down at $x = 2$.
- We are given that $f(2) = -3$. Since f is increasing at $x = 2$, we expect $f(2.1)$ to be greater than $f(2) = -3$.
- Since $f''(2)$ is negative, we also know that f' is decreasing at 2. Thus, we expect that $f'(2.1)$ will be less than $f'(2) = 1.5$.
- In the following figure, we see a possible graph of a function f that passes through the point $(2, -3)$ at an instantaneous rate of $f'(2) = 1.5$ and in such a way that its second derivative at $x = 2$ is negative.



7. For a certain function $y = g(x)$, its derivative is given by the function pictured in Figure 1.7.19.

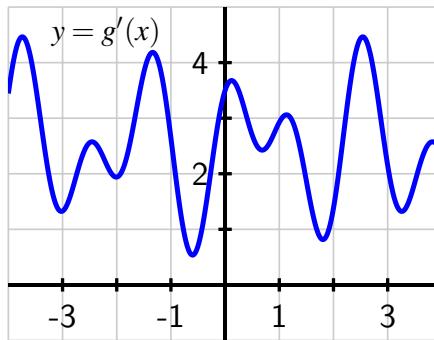


Figure 1.7.19: The graph of $y = g'(x)$.

- What is the approximate slope of the tangent line to $y = g(x)$ at the point $(2, g(2))$?
- 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' .
- On the interval $-3 < x < 3$, how many times does the concavity of g change? Why?
- Use the provided graph to estimate the value of $g''(2)$.

Answer.

- $g'(2) \approx 1.4$.
- At most one.
- 9.
- $g''(2) \approx 5.5$.

Solution.

- From the given graph of g' , we may estimate that $g'(2) \approx 1.4$.
- Observe that g' is always positive. This tells us that g must be always increasing. Therefore, it follows that g may cross the x -axis at most one time, and hence there can be at most one solution to $g(x) = 0$.
- On $-3 < x < 3$, g changes concavity g times, since g' changes from increasing to decreasing or from decreasing to increasing 9 times. Whenever g' is increasing, g'' is positive, and whenever g' is decreasing, g'' is negative. Hence, whenever g' changes from increasing to decreasing or vice versa, it follows that g'' changes sign, and this causes a change in the concavity of g . We therefore see that while g is always increasing, g changes concavity many times.
- From the given graph of g' , we can see that $g'(1.9) \approx 0.9$ and $g'(2.1) \approx 2$. Using a central difference, it follows that

$$g''(2) \approx \frac{g'(2.1) - g'(1.9)}{0.2} \approx \frac{2 - 0.9}{0.2} = 5.5.$$

- A bungee jumper's height h (in feet) at time t (in seconds) is given in part by the table:

t	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$h(t)$	200	184.2	159.9	131.9	104.7	81.8	65.5	56.8	55.5	60.4	69.8
t	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	
$h(t)$	81.6	93.7	104.4	112.6	117.7	119.4	118.2	114.8	110.0	104.7	

- 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?
- Use the given data and your work in (a) to estimate $h''(5)$.
- What physical property of the bungee jumper does the value of $h''(5)$ measure? What are its units?
- 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?

Answer.

- $h'(4.5) \approx 14.3$; $h'(5) \approx 21.2$; $h'(5.5) \approx 23.9$; rising most rapidly at $t = 5.5$.
- $h'(5) \approx 9.6$.
- Acceleration of the bungee jumper in feet per second per second.
- $0 < t < 2$, $6 < t < 10$.

Solution.

- We will use central differences to estimate the requested derivatives. Thus,

$$h'(4.5) \approx \frac{h(5) - h(4)}{5 - 4} = \frac{69.8 - 55.5}{1} = 14.3$$

$$h'(5) \approx \frac{h(5.5) - h(4.5)}{5.5 - 4.5} = \frac{81.6 - 60.4}{1} = 21.2$$

$$h'(5.5) \approx \frac{h(6) - h(5)}{6 - 5} = \frac{93.7 - 69.8}{1} = 23.9$$

The bungee jumper is rising most rapidly at $t = 5.5$, at a rate of about 23.9 feet per second.

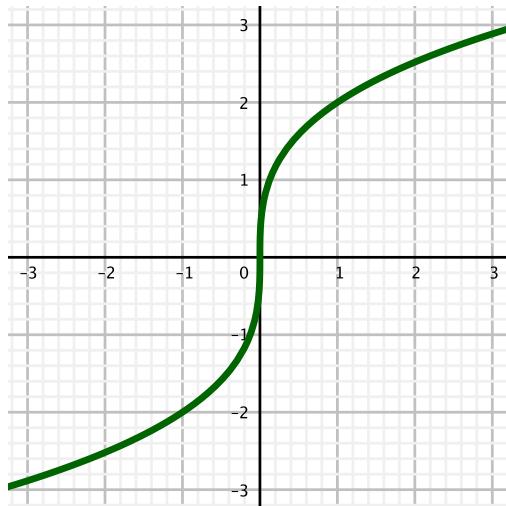
- b. We again use a central difference, this time to estimate $h''(5)$. Observe that

$$h''(5) \approx \frac{h'(5.5) - h'(4.5)}{5.5 - 4.5} = \frac{23.9 - 14.3}{1} = 9.6$$

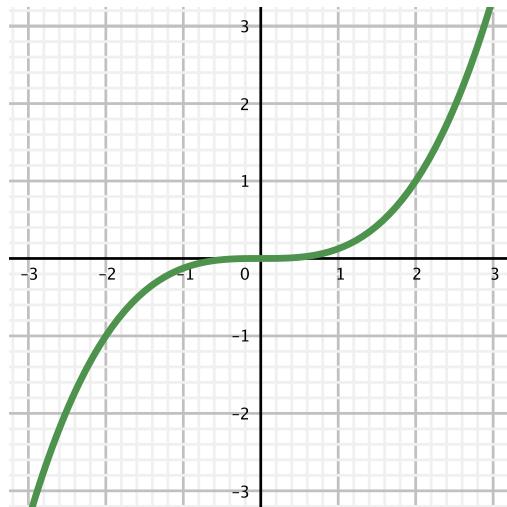
- c. $h''(5)$ measures the acceleration of the bungee jumper at $t = 5$. Thus we know that the person's acceleration at $t = 5$ is about 9.6 feet per second per second. This measures the instantaneous rate of change of the jumper's velocity.
- d. The bungee jumper's position function is concave down on approximately the intervals $0 < t < 2, 6 < t < 10$. This is easiest to see if we plot the data using a program like Excel. On these time intervals, the velocity of the jumper is decreasing: either the velocity is positive and getting less positive, or negative and getting more negative.
9. For each prompt that follows, sketch a possible graph of a function on the interval $-3 < x < 3$ that satisfies the stated properties.
- $y = f(x)$ such that f is increasing on $-3 < x < 3$, concave up on $-3 < x < 0$, and concave down on $0 < x < 3$.
 - $y = g(x)$ such that g is increasing on $-3 < x < 3$, concave down on $-3 < x < 0$, and concave up on $0 < x < 3$.
 - $y = h(x)$ such that h is decreasing on $-3 < x < 3$, concave up on $-3 < x < -1$, neither concave up nor concave down on $-1 < x < 1$, and concave down on $1 < x < 3$.
 - $y = p(x)$ such that p is decreasing and concave down on $-3 < x < 0$ and is increasing and concave down on $0 < x < 3$.

Answer.

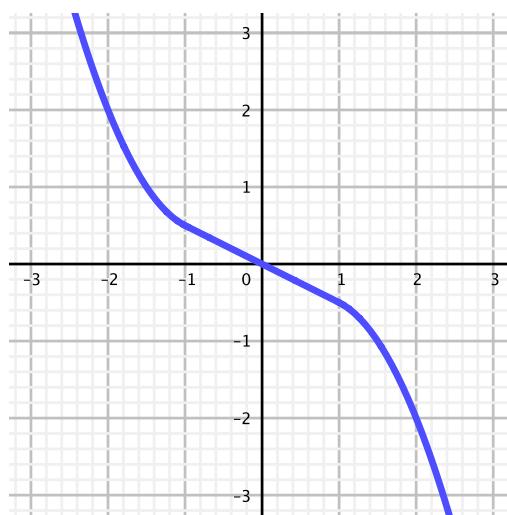
a.



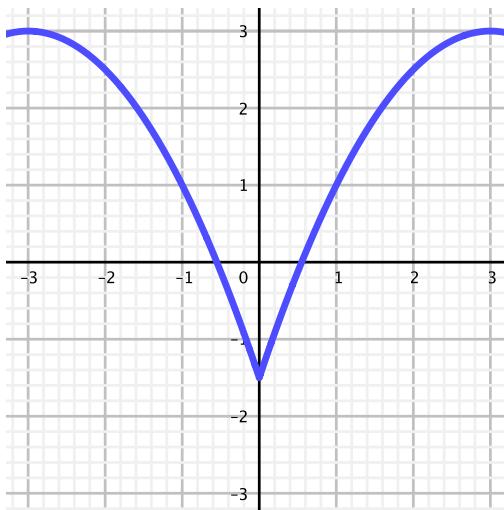
b.



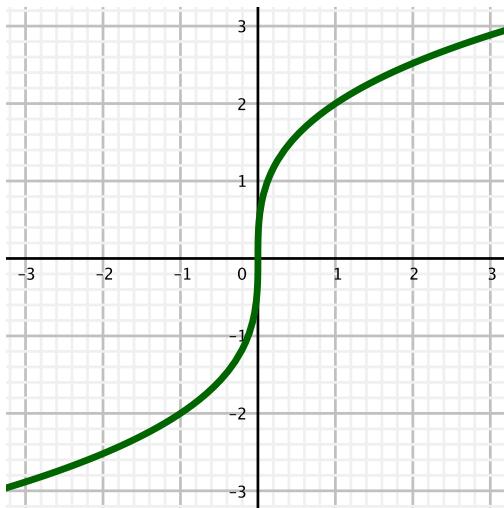
c.



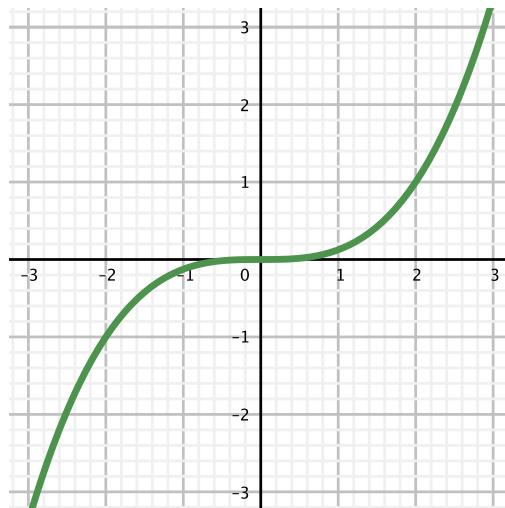
d.

**Solution.**

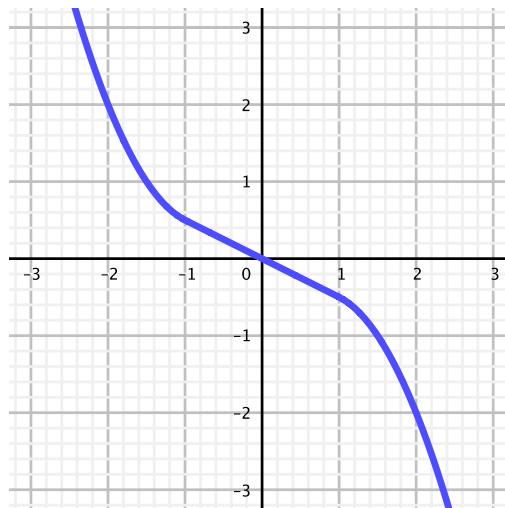
- a. When a function is increasing, its graph rises as x increases, and when a function is concave up, that means the graph of the function looks bowl-shaped. A graph of an increasing, concave up function f on the interval $(-3, 0)$ is shown in the following figure. When a function is concave down, its graph looks like an upside down bowl. The function f in the figure is increasing and concave down on the interval $(0, 3)$. Note that f is also increasing on the entire interval $(-3, 3)$.



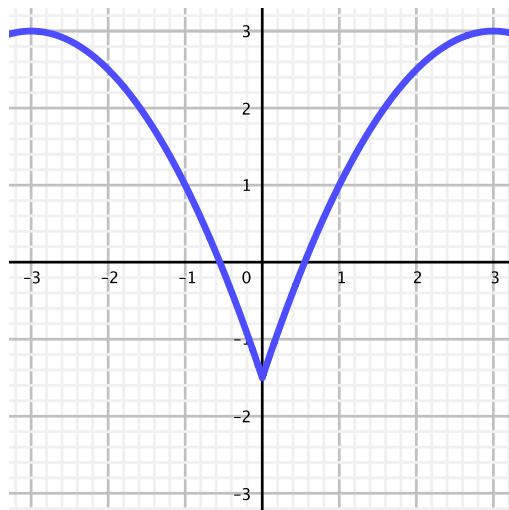
- b. When a function is increasing, its graph rises as x increases, and when a function is concave down, that means the graph of the function looks like an upside down bowl. A graph of an increasing, concave down function g on the interval $(-3, 0)$ is shown in the figure below. When a function is concave up, its graph looks like a bowl. The function g in the figure is increasing and concave up on the interval $(0, 3)$. Note that g is also increasing on the entire interval $(-3, 3)$.



- c. When a function is decreasing, its graph falls as x increases, and when a function is concave up, that means the graph of the function looks bowl-shaped. A graph of a decreasing, concave up function h on the interval $(-3, -1)$ is shown in the figure below. To be neither concave up or concave down, a graph has to be linear, so the graph of h in the following figure is a straight line on $(-1, 1)$. When a function is concave down, its graph looks like an upside down bowl. The function h in the figure is decreasing and concave down on the interval $(1, 3)$. Note that h is decreasing on the entire interval $(-3, 3)$.



- d. When a function is decreasing, its graph falls as x increases, and when a function is concave down, that means the graph of the function looks like an upside down bowl. A graph of a decreasing, concave down function p on the interval $(-3, 0)$ is shown in the figure below. When a function is increasing, its graph rises as x increases. The function p in the figure is increasing and concave down on the interval $(0, 3)$.



1.8 Differentiability

Motivating Questions

- 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?
 - How are the characteristics of a function having a limit, being continuous, and being differentiable at a given point related to one another?
-

In this section we aim to determine whether or not the function has a derivative $f'(a)$ at $x = a$.

1.8.1 Being differentiable at a point

We recall that a function f is said to be differentiable at $x = a$ if $f'(a)$ exists. Moreover, for $f'(a)$ to exist, we know that the graph of $y = f(x)$ must have a tangent line at the point $(a, f(a))$, since the value of $f'(a)$ is precisely the slope of this line. Observe that in order to ask if f has a tangent line at $(a, f(a))$, it is necessary for f to 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 $\lim_{x \rightarrow a} f(x)$, then it doesn't 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 $x = a$. Stated differently, if f is not continuous at $x = a$, then it is automatically the case that f is not differentiable there. For example, in Figure 1.2.15, both f and g fail to be differentiable at $x = 1$ because neither function is continuous at $x = 1$.

A natural question to ask at this point is “is there a difference between continuity and differentiability?” Can a function fail to be differentiable at a point where the function is continuous? To answer these questions, we consider a certain function f , where the graph of $y = f(x)$ is displayed below in Figure 1.8.1. We notice that f has a sharp corner at the point $(1, 1)$, and further observe that f is continuous at $x = 1$ since $\lim_{x \rightarrow 1} f(x) = 1 = f(1)$.

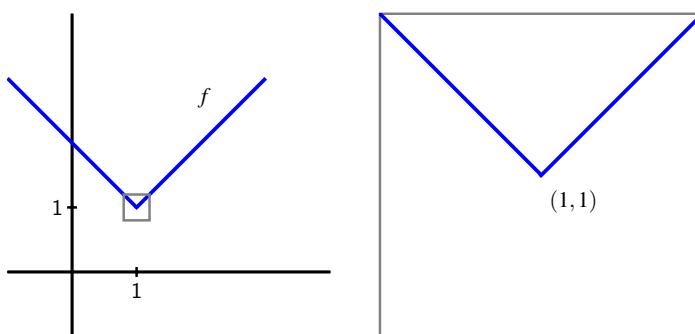


Figure 1.8.1: A function f that is continuous at $x = 1$ but not differentiable at $x = 1$; at right, we zoom in on the point $(1, 1)$ in a magnified version of the box in the left-hand plot.

However, the function f in Figure 1.8.1 is not differentiable at $x = 1$ 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 for the value of the derivative at $x = 1$. Alternatively, 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 see visually in Figure 1.8.1 that this function does not have a tangent line at $x = 1$. Regardless of how closely we examine the function by zooming in on $(1, 1)$ on the graph of $y = f(x)$, 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.

If a function *does* have a tangent line at a given point, then the function and the tangent line should appear essentially indistinguishable when we zoom in on the point of tangency¹. Conversely, if we zoom in on a point and the function looks like a single straight line, then the function should have a tangent line there, and thus be differentiable at that point. Therefore, we say that a function that is differentiable at $x = a$ is *locally linear*.

To summarize the preceding discussion of differentiability, we make several important observations.

- If f is differentiable at $x = a$, then f is continuous at $x = a$. Equivalently, if f fails to be continuous at $x = a$, then f will not be differentiable at $x = a$.
- A function can be continuous at a point without being differentiable there. In particular, a function f is not differentiable at $x = a$ if the graph has a sharp corner (or *cusp*) at the point $(a, f(a))$.
- If f is differentiable at $x = a$, then f is locally linear at $x = a$. In other words, a differentiable function looks linear when viewed up close because it resembles its tangent line at any given point of differentiability.

¹For instance, see <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 1.8.2. In this example, we explore two different functions and classify the points at which each is not differentiable. Let g be the function given by the rule $g(x) = |x|$, and let f be the function whose graph is given below in Figure 1.8.3.

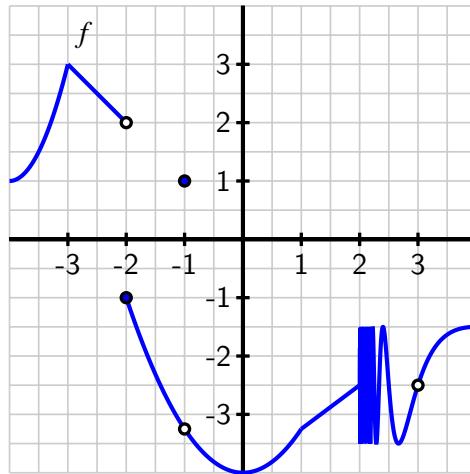


Figure 1.8.3: The graph of $y = f(x)$ for Example 1.8.2.

- Visualize (or sketch) the graph of $y = g(x)$, and use it to explain why g is differentiable at every nonzero value of x .
- Use the limit definition of the derivative to show that $g'(0) = \lim_{h \rightarrow 0} \frac{|h|}{h}$. Then use small positive and negative values of h to explain why this limit does not exist. Conclude that g is not differentiable at $x = 0$.
- State all values of a for which f is not differentiable at $x = a$. For each, provide a reason for your conclusion.
- True or false: if a function p is differentiable at $x = b$, then $\lim_{x \rightarrow b} p(x)$ must exist. Write at least one sentence to justify your choice.

Hint.

- What type of function is g for all $x < 0$? For all $x > 0$?
- Recall that $g'(0) = \lim_{h \rightarrow 0} \frac{g(0+h) - g(0)}{h}$. What is the value of $|h|$ when $h < 0$?
- You might start by identifying points where f is not continuous.
- What does being differentiable at a point tell you about continuity there?

Answer.

a. g is piecewise linear.

b. Show that

$$g'(0) = \lim_{h \rightarrow 0} \frac{|h|}{h},$$

which does not exist.

c. $a = -3, -2, -1, 1, 2, 3$.

d. True.

Solution.

a. We know that $g(x) = |x|$ is given by the formula $g(x) = -x$ when $x < 0$ and by $g(x) = x$ when $x \geq 0$. Each of these pieces of g is a straight line, so at every point other than the point where they meet, the function g has a well-defined slope, and thus is differentiable.

b. Observe that

$$\begin{aligned} g'(0) &= \lim_{h \rightarrow 0} \frac{g(0+h) - g(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{|0+h| - |0|}{h} \\ &= \lim_{h \rightarrow 0} \frac{|h|}{h} \end{aligned}$$

To evaluate this limit, note that $|h| = h$ whenever $h > 0$, and thus

$$\lim_{h \rightarrow 0^+} \frac{|h|}{h} = \lim_{h \rightarrow 0^+} \frac{h}{h} = 1.$$

In contrast, $|h| = -h$ whenever $h < 0$, and thus

$$\lim_{h \rightarrow 0^-} \frac{|h|}{h} = \lim_{h \rightarrow 0^-} \frac{-h}{h} = -1$$

Since the right- and left-hand limits are not equal, it follows that

$$g'(0) = \lim_{h \rightarrow 0} \frac{|h|}{h}$$

does not exist.

- c. f is not differentiable at $a = -2, -1, 2, 3$ because at each of these points f is not continuous. In addition, f is not differentiable at $a = -3$ and $a = 1$ because the graph of f has a corner point (or cusp) at each of these values.
- d. We know that a function f is continuous whenever it is differentiable, and that one characteristic of f being continuous at $x = a$ is that $\lim_{x \rightarrow a} f(x)$ exists. Therefore the statement “if a function p is differentiable at $x = b$, then $\lim_{x \rightarrow b} p(x)$ must exist” is true.

1.8.2 Summary

- A function f is differentiable at $x = a$ whenever $f'(a)$ exists, which means that f has a tangent line at $(a, f(a))$ and thus f is locally linear at $x = a$. Informally, this means that the function looks like a line when viewed up close at $(a, f(a))$ and that there is not a corner point or cusp at $(a, f(a))$.
- Differentiability is a stronger condition than continuity, which is a stronger condition than having a limit. In particular, if f is differentiable at $x = a$, then f is also continuous at $x = a$, and if f is continuous at $x = a$, then f has a limit at $x = a$.

1.8.3 Exercises

1. (Continuity and differentiability of a graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. Consider the graph of the function $y = p(x)$ that is provided in Figure 1.8.4. Assume that each portion of the graph of p is a straight line, as pictured.

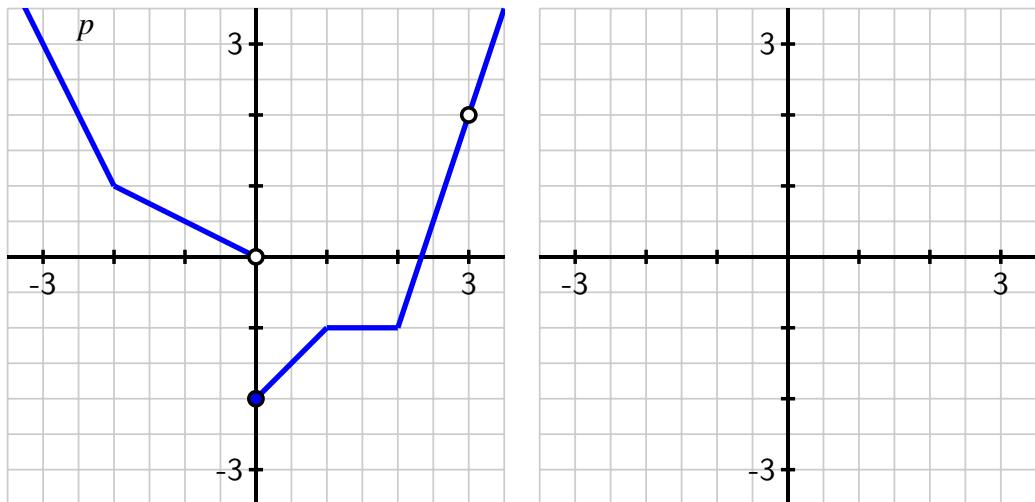
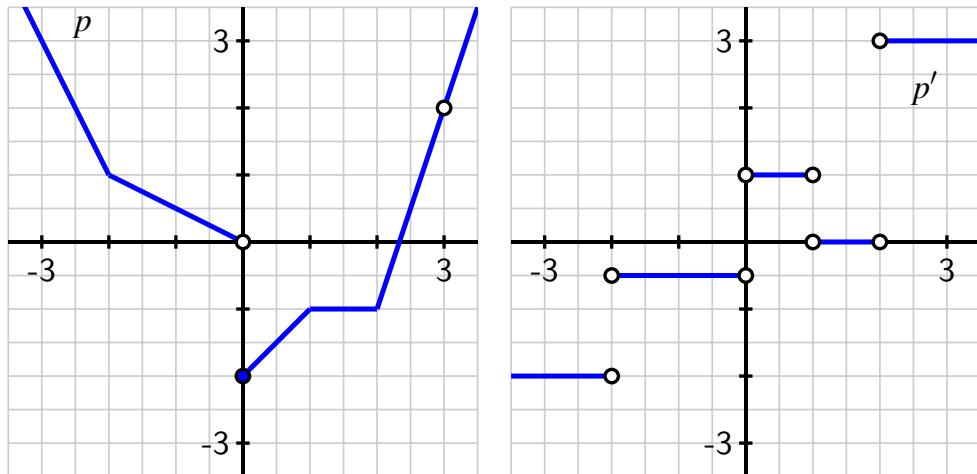


Figure 1.8.4: At left, the piecewise linear function $y = p(x)$. At right, axes for plotting $y = p'(x)$.

- a. State all values of a for which $\lim_{x \rightarrow a} p(x)$ 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 in Figure 1.8.4, sketch an accurate graph of $y = p'(x)$.

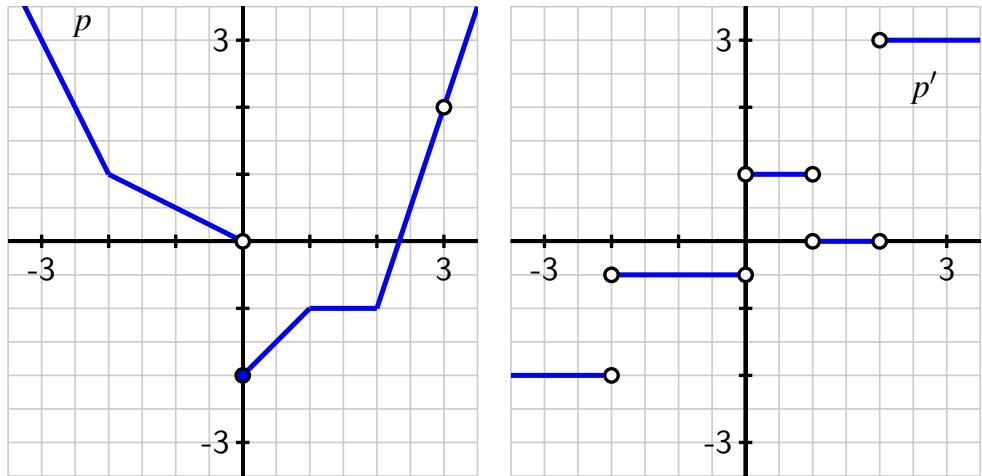
Answer.

- a. $a = 0$.
- b. $a = 0, 3$.
- c. $a = -2, 0, 1, 2, 3$.
- d.



Solution.

- a. The only value of a at which $\lim_{x \rightarrow a} p(x)$ does not exist is $a = 0$. This is because $\lim_{x \rightarrow 0^-} p(x) = 0 \neq -2 = \lim_{x \rightarrow 0^+} p(x)$.
- b. p is not continuous at $a = 0$ because $\lim_{x \rightarrow 0} p(x)$ does not exist. In addition, p is not continuous at $a = 3$ because $p(3)$ is not defined.
- c. p is not differentiable at $a = 0$ and at $a = 3$ since p is not continuous at these points. In addition, p is not differentiable at $a = -2, 1, 2$ since p has corner points at these values, making p not have a tangent line at these values.
- d. Note that p is piecewise linear, so the value of the derivative of p (where it exists) is the slope of p , which will be constant on each interval where p is linear. In addition, note that p' is not defined at $a = -2, 0, 1, 2, 3$, which is reflected by open circles in the following plot.

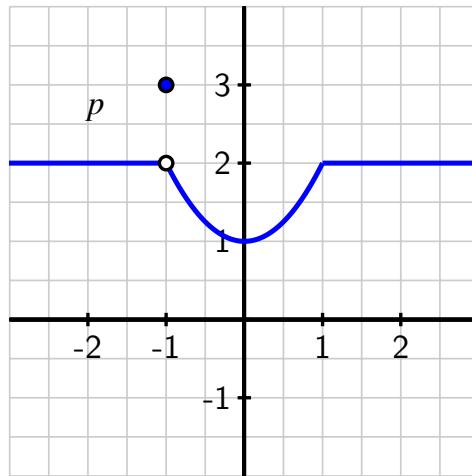


3. For each of the following prompts, give an example of a function that satisfies the stated criteria. A formula or a graph, with reasoning, is sufficient for each. If no such example is possible, explain why.

- A function f that is continuous at $a = 2$ but not differentiable at $a = 2$.
- A function g that is differentiable at $a = 3$ but does not have a limit at $a = 3$.
- A function h that has a limit at $a = -2$, is defined at $a = -2$, but is not continuous at $a = -2$.
- A function p that satisfies all of the following:
 - $p(-1) = 3$ and $\lim_{x \rightarrow -1} p(x) = 2$
 - $p(0) = 1$ and $p'(0) = 0$
 - $\lim_{x \rightarrow 1} p(x) = p(1)$ and $p'(1)$ does not exist

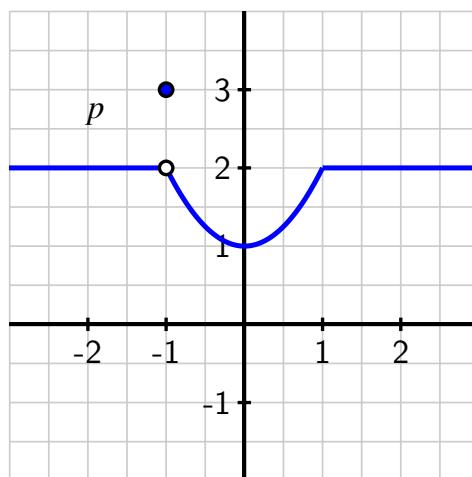
Answer.

- $f(x) = |x - 2|$.
- Impossible.
- Let f be the function defined to be $f(x) = 1$ for every value of $x \neq -2$, and such that $f(-2) = 4$.
-



Solution.

- The function $f(x) = |x - 2|$ is continuous at $a = 2$, because $f(2) = 0 = \lim_{x \rightarrow 2} f(x)$, but not differentiable at $a = 2$ because f has a corner point at $a = 2$.
- This is impossible. Any function that is differentiable at a point must have a limit at the point.
- Let f be the function defined to be $f(x) = 1$ for every value of $x \neq -2$, and such that $f(-2) = 4$. This function has a limit of 1 as $x \rightarrow -2$, but since $f(-2) = 4 \neq 1$, the function is not continuous at $a = -2$.
- A possible function p that has the desired properties is shown in the following figure.



- Consider the function $g(x) = \sqrt{|x|}$.

- Use a graph to explain visually why g is not differentiable at $x = 0$.

- b. Use the limit definition of the derivative to show that

$$g'(0) = \lim_{h \rightarrow 0} \frac{\sqrt{|h|}}{h}.$$

- c. Investigate the value of $g'(0)$ by estimating the limit in (b) using small positive and negative values of h . For instance, you might compute $\frac{\sqrt{|-0.01|}}{0.01}$. Be sure to use several different values of h (both positive and negative), including ones closer to 0 than 0.01. What do your results tell you about $g'(0)$?
- d. Use your graph in (a) to sketch an approximate graph of $y = g'(x)$.

Answer.

- a. At $x = 0$.

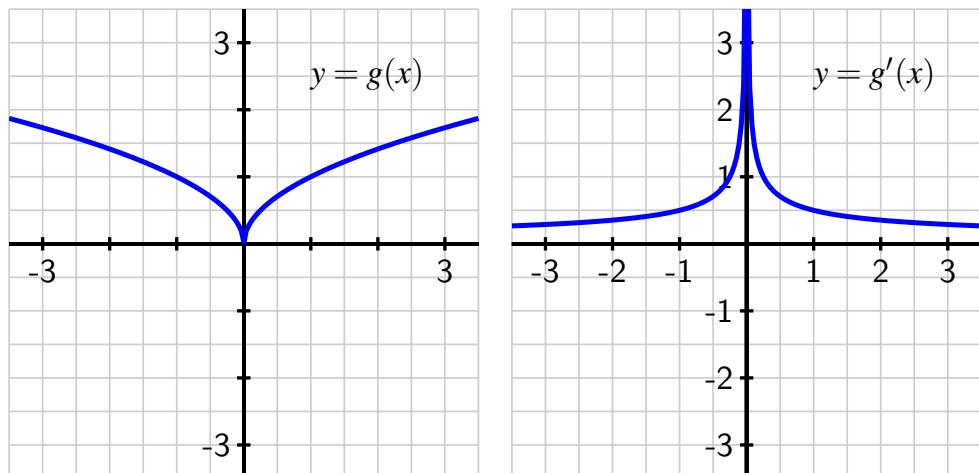
$$\begin{aligned} g'(0) &= \lim_{h \rightarrow 0} \frac{g(0 + h) - g(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{|h|} - \sqrt{|0|}}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{|h|}}{h} \end{aligned}$$

- b.

h	0.1	0.01	0.001	0.0001	0.00001	-0.1	-0.01	-0.001	-0.0001
$\sqrt{ h }/h$	3.162	10	31.62	100	-3.162	-10	-31.62	-100	

$g'(0)$ does not exist.

- c.



Solution.

- a. As seen at left in the figure below in part (d), g has a sharp corner point at $x = 0$, which makes g not differentiable at this point.

By definition, we know

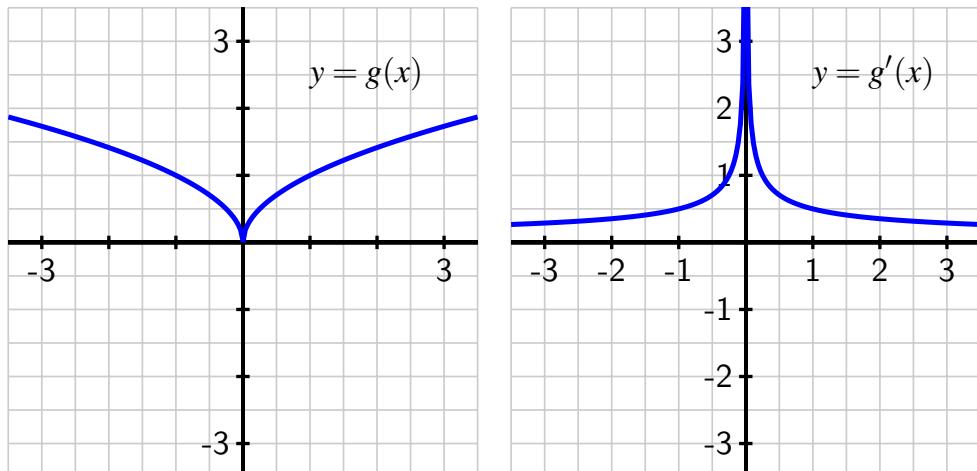
$$\begin{aligned} g'(0) &= \lim_{h \rightarrow 0} \frac{g(0+h) - g(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{|h|} - \sqrt{|0|}}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sqrt{|h|}}{h} \end{aligned}$$

- b. In the following table, we record the values of $\frac{\sqrt{|h|}}{h}$ for small values of h .

h	0.1	0.01	0.001	0.0001	-0.1	-0.01	-0.001	-0.0001
$\sqrt{ h }/h$	3.162	10	31.62	100	-3.162	-10	-31.62	-100

These results show from a different perspective that $g'(0)$ does not exist. For small negative h values, the difference quotient is negative and getting more negative without bound. For small positive values of h , the difference quotient is positive and increasing without bound. This shows that $\lim_{h \rightarrow 0} \frac{\sqrt{|h|}}{h}$ does not exist.

- c. An approximate graph of g' is shown at right in the figure below.



Chapter 2

Computing Derivatives

2.1 Elementary derivative rules

Motivating Questions

- What are alternate notations for the derivative?
 - How can we use the algebraic structure of a function $f(x)$ to 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 the form $f(x) = a^x$?
 - If we know the derivative of $y = f(x)$, what is the derivative of $y = kf(x)$, where k is a constant?
 - If we know the derivatives of $y = f(x)$ and $y = g(x)$, how do we compute the derivative of $y = f(x) + g(x)$?
-

In Chapter 1 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 . The derivative also tells us the slope of the tangent line to the graph of $y = f(x)$ at any given value of x . So far, we have focused on interpreting the derivative graphically or, in the context of a physical setting, as a meaningful rate of change. To calculate the value of the derivative at a specific point, we have relied on the limit definition of the derivative:

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

In this chapter we investigate how the limit definition of the derivative leads to interesting patterns and rules that enable us to quickly find a formula for $f'(x)$ *without* directly using

the limit definition. For example, we would like to apply shortcuts to differentiate a function such as $g(x) = 4x^7 - \sin(x) + 3e^x$

Example 2.1.1. Functions of the form $f(x) = x^n$, where $n = 1, 2, 3, \dots$, are often called **power functions**. The first two questions below revisit work we did earlier in Chapter 1, 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 $f'(x)$?

Hint.

- a $(a + b)^2 = a^2 + 2ab + b^2$
- b $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$
- c $(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$
- d Use your answers to (a), (b), and (c) to find a pattern in the relationship between the formula for f and the formula for f' in each case.
- e What pattern did you use in (d) to find $f'(x)$ when $f(x) = x^5$? What happens when you replace the “5” with “ n ”?

Answer.

- a $f'(x) = 2x$.
- b $f'(x) = 3x^2$.
- c $f'(x) = 4x^3$.
- d $f'(x) = 5x^4$ when $f(x) = x^5$; when $f(x) = x^{13}$, then $f'(x) = 13x^{12}$.
- e $f'(x) = nx^{n-1}$, where $f(x) = x^n$ and n is a positive integer.

Solution.

a

$$\begin{aligned}
 f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x^2 + 2xh + h^2) - x^2}{h} \\
 &= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} \\
 &= \lim_{h \rightarrow 0} (2x + 2h) \\
 &= 2x.
 \end{aligned}$$

b

$$\begin{aligned}
 f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x+h)^3 - x^3}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x^3 + 3x^2h + 3xh^2 + h^3) - x^3}{h} \\
 &= \lim_{h \rightarrow 0} \frac{3x^2h + 3xh^2 + h^3}{h} \\
 &= \lim_{h \rightarrow 0} (3x^2 + 3xh + h^2) \\
 &= 3x^2
 \end{aligned}$$

c

$$\begin{aligned}
 f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x+h)^4 - x^4}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x^4 + 4x^3h + 6x^2h^2 + 4xh^3 + h^4) - x^4}{h} \\
 &= \lim_{h \rightarrow 0} \frac{4x^3h + 6x^2h^2 + 4xh^3 + h^4}{h} \\
 &= \lim_{h \rightarrow 0} (4x^3 + 6x^2h + 4xh^2 + h^3) \\
 &= 4x^3
 \end{aligned}$$

- d In the previous parts, we discovered that the derivatives of x^2 , x^3 , and x^4 are $2x$, $3x^2$, and $4x^3$, respectively. It looks like when we increase the power of x by one, this affects the derivative by increasing both the coefficient and the power of x by one. For

instance, going from $f(x) = x^2$ to $g(x) = x^3$, we saw the derivatives go from $f'(x) = 2x$ to $g'(x) = 3x^2$.

Since the derivative of x^4 is $4x^3$, if this pattern continues, then it makes sense for the derivative of x^5 to be $5x^4$. Then if $f(x) = x^{13}$, we've increased the power of x by eight from x^5 , so we want to increase the coefficient and power in the derivative by eight as well, to get $f'(x) = (5 + 8)x^{4+8} = 13x^{12}$.

- e The pattern we've seen is that increasing the power of x by one makes the coefficient and power of x in the derivative by one. Also, the power of x in the derivative is one less than in the original function in each case we've looked at, and the coefficient on the derivative is the same as the power in the original function. For instance, $f(x) = x^5$ implies $f'(x) = 5x^4 = 5x^{5-1}$. If we try to generalize this by replacing "5" with " n ", we conjecture that when n is a positive integer and $f(x) = x^n$, then $f'(x) = nx^{n-1}$.

2.1.1 Some Key Notation

In addition to our usual f' notation, there are other ways to denote the derivative of a function, as well as the instruction to take the derivative. If we are thinking about the relationship between y and x , we sometimes denote the derivative of y with respect to x by the symbol

$$\frac{dy}{dx}$$

which we read "dee-y dee-x." For example, if $y = x^2$, we'll write that the derivative is $\frac{dy}{dx} = 2x$. This notation comes from the fact that the derivative is related to the slope of a line, and slope is measured by $\frac{\Delta y}{\Delta x}$. Note that while we read $\frac{\Delta y}{\Delta x}$ as "change in y over change in x ," we view $\frac{dy}{dx}$ as a single symbol, not a quotient of two quantities.

We use a variant of this notation as the instruction to take the derivative. In particular,

$$\frac{d}{dx} [\square]$$

means "take the derivative of the quantity in \square with respect to x ." For example, we may write $\frac{d}{dx}[x^2] = 2x$.

It is important to note that the independent variable can be different from x . If we have $f(z) = z^2$, we then write $f'(z) = 2z$. Similarly, if $y = t^2$, we say $\frac{dy}{dt} = 2t$. And it is also true that $\frac{d}{dq}[q^2] = 2q$. This notation may also be used for second derivatives: $f''(z) = \frac{d}{dz} \left[\frac{df}{dz} \right] = \frac{d^2 f}{dz^2}$.

In what follows, we'll build a repertoire of functions for which we can quickly compute the derivative.

2.1.2 Constant, Power, and Exponential Functions

So far, we know the derivative formula for two important classes of functions: constant functions and power functions. If $f(x) = c$ is a constant function, 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 Functions

For any real number c , if $f(x) = c$, then $f'(x) = 0$.

Example 2.1.2. If $f(x) = 7$, then $f'(x) = 0$. Similarly, $\frac{d}{dx}[\sqrt{3}] = 0$.

In your work in Example 2.1.1, you conjectured that for any positive integer n , if $f(x) = x^n$, then $f'(x) = nx^{n-1}$. This rule can be formally proved for any positive integer n , and also for any nonzero real number (positive or negative).

Power Functions

For any nonzero real number n , if $f(x) = x^n$, then $f'(x) = nx^{n-1}$.

Example 2.1.3. Using the rule for power functions, we can compute the following derivatives. If $g(z) = z^{-3}$, then $g'(z) = -3z^{-4}$. Similarly, if $h(t) = t^{7/5}$, then $\frac{dh}{dt} = \frac{7}{5}t^{2/5}$, and $\frac{d}{dq}[q^\pi] = \pi q^{\pi-1}$.

It will be instructive to have a derivative formula for one more type of basic function. For now, we simply state this rule without explanation or justification; we will explore why this rule is true in one of the exercises. And we will encounter graphical reasoning for why the rule is plausible in Example 2.3.1.

Exponential Functions

For any positive real number a , if $f(x) = a^x$, then $f'(x) = a^x \ln(a)$.

Example 2.1.4. 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!

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.

Example 2.1.5. 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. $f(t) = \pi$ | d. $p(x) = 3^{1/2}$ | f. $s(q) = q^{-1}$ |
| b. $g(z) = 7^z$ | | |
| c. $h(w) = w^{3/4}$ | e. $r(t) = (\sqrt{2})^t$ | g. $m(t) = \frac{1}{t^3}$ |

Hint.

- a. Is π a variable or a constant?
- b. Is g a power or exponential function?
- c. Is h a power or exponential function?
- d. Is $3^{1/2}$ a constant or a variable?
- e. $\sqrt{2}$ is a constant
- f. Remember the notation here means “take the derivative with respect to q of q^{-1} .”
- g. Rewrite the fraction using a negative exponent.

Answer.

- a. $f'(t) = 0$.
- b. $g'(z) = 7^z \ln(7)$.
- c. $h'(w) = \frac{3}{4}w^{-1/4}$.
- d. $\frac{dp}{dx} = 0$.
- e. $r'(t) = (\sqrt{2})^t \ln(\sqrt{2})$.
- f. $\frac{d}{dq}[q^{-1}] = -q^{-2}$.
- g. $\frac{dm}{dt} = -3t^{-4} = -\frac{3}{t^4}$.

Solution.

- a. $f(t) = \pi$ is constant, so $f'(t) = 0$.
- b. $g(z) = 7^z$ is an exponential function, so $g'(z) = 7^z \ln(7)$.
- c. $h(w) = w^{3/4}$ is a power function, thus $h'(w) = \frac{3}{4}w^{-1/4}$.
- d. $p(x) = 3^{1/2}$ is constant, and therefore $\frac{dp}{dx} = 0$.
- e. $r(t) = (\sqrt{2})^t$ is exponential (since $\sqrt{2}$ is a constant), and so we have $r'(t) = (\sqrt{2})^t \ln(\sqrt{2})$.
- f. $\frac{d}{dq}[q^{-1}] = -q^{-2}$, by the rule for power functions.
- g. $m(t) = \frac{1}{t^3} = t^{-3}$, so $\frac{dm}{dt} = -3t^{-4} = -\frac{3}{t^4}$.

2.1.3 Constant Multiples and Sums of Functions

Next we will learn how to compute the derivative of a function constructed as an algebraic combination of basic functions. For instance, we'd like to be able to take the derivative of a polynomial function such as

$$p(t) = 3t^5 - 7t^4 + t^2 - 9,$$

which is a sum of constant multiples of powers of t . To that end, we develop two new rules: the Constant Multiple Rule and the Sum Rule.

How is the derivative of $y = kf(x)$ related to the derivative of $y = f(x)$? 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, so the slope of the function $y = kf(x)$ is k times as steep as the slope of $y = f(x)$. Thus, when we multiply a function by a factor of k , we change the value of its derivative by a factor of k as well.¹

The Constant Multiple Rule

For any real number k , if $f(x)$ is a differentiable function with derivative $f'(x)$, then $\frac{d}{dx}[kf(x)] = kf'(x)$.

In words, this rule says that “the derivative of a constant times a function is the constant times the derivative of the function.”

Example 2.1.6. 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 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 is the value of the new function the sum of the values of the two known functions, but the slope of the new function is the sum of the slopes of the known functions. Therefore², 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 $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$.

In words, the Sum Rule tells us that “the derivative of a sum is the sum of the derivatives.” It also tells us that a sum of two differentiable functions is also 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.” We can now compute derivatives of sums and differences of elementary functions.

Example 2.1.7. Using the sum rule, $\frac{d}{dw}(2^w + w^2) = 2^w \ln(2) + 2w$. Using both the sum and constant multiple rules, if $h(q) = 3q^6 - 4q^{-3}$, then $h'(q) = 3(6q^5) - 4(-3q^{-4}) = 18q^5 + 12q^{-4}$.

¹The Constant Multiple Rule can be formally proved as a consequence of properties of limits, using the limit definition of the derivative.

²Like 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.

Example 2.1.8. 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. $q(x) = \frac{x^3 - x + 2}{x}$
 g. $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$

Hint.

- a. Use the sum rule.
- b. Use the sum rule together with the constant multiple rule.
- c. How can you rewrite \sqrt{z} using exponents?
- d. Is e^4 a constant or variable?
- e. Expand the product before attempting to find the derivative.
- f. Rewrite the single fraction as a sum of three fractions, and simplify.
- g. Note that “ a ” is the independent variable.

Answer.

a. $f'(x) = \frac{5}{3}x^{2/3} - 4x^3 + 2^x \ln(2).$
 b. $g'(x) = 14e^x + 3 \cdot 5x^4 - 1.$
 c. $h'(z) = \frac{1}{2}z^{-1/2} - 4z^{-5} + 5^z \ln(5).$
 d. $\frac{dr}{dt} = \sqrt{53} \cdot 7t^6 - \pi e^t.$
 e. $\frac{ds}{dy} = 4y^3.$
 f. $q'(x) = 2x - 2x^{-2}.$
 g. $p'(a) = 12a^3 - 6a^2 + 14a - 1.$

Solution.

a. $f(x) = x^{5/3} - x^4 + 2^x$, so by the sum rule, $f'(x) = \frac{5}{3}x^{2/3} - 4x^3 + 2^x \ln(2).$

- b. $g(x) = 14e^x + 3x^5 - x$, so by the sum and constant multiple rules, $g'(x) = 14e^x + 3 \cdot 5x^4 - 1$.
- c. $h(z) = \sqrt{z} + \frac{1}{z^4} + 5^z = z^{1/2} + z^{-4} + 5^z$, thus $h'(z) = \frac{1}{2}z^{-1/2} - 4z^{-5} + 5^z \ln(5)$.
- d. Since $r(t) = \sqrt{53}t^7 - \pi e^t + e^4$ and $\sqrt{53}$, π , and e^4 are all constants, it follows from the sum and constant multiple rules, as well as the derivative of a constant rule, that $\frac{dr}{dt} = \sqrt{53} \cdot 7t^6 - \pi e^t$. (Note particularly that $\frac{d}{dt}[e^4] = 0$ since e^4 is constant.)
- e. $s(y) = (y^2 + 1)(y^2 - 1) = y^4 - 1$, thus $\frac{ds}{dy} = 4y^3$.
- f. $q(x) = \frac{x^3 - x + 2}{x} = \frac{x^3}{x} - \frac{x}{x} + \frac{2}{x} = x^2 - 1 + 2x^{-1}$. Now it follows that $q'(x) = 2x - 2x^{-2}$.
- g. $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$, so $p'(a) = 12a^3 - 6a^2 + 14a - 1$.

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 .” Similarly, if the derivative exists at a point, we say “ f is differentiable at that point,” or that f can be differentiated.

As we work with the algebraic structure of functions, it is important to develop a big picture view of what we are doing. Here, we make several general observations based on the rules we have so far.

- 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.
- 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.
- We should not lose sight of the fact that all of the meaning of the derivative that we developed in Chapter 1 still holds. The 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.

Example 2.1.9. 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.

- Find the slope of the tangent line to $h(z) = \sqrt{z} + \frac{1}{z}$ at the point where $z = 4$.
- 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.
 - Determine the instantaneous rate at which the population is growing on day 4, and include units in your answer.
 - 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))?

Hint.

- a. How would $h'(z)$ help you answer the question?
- b. Think about finding both $P'(t)$ and $P''(t)$.
- c. What two important pieces of information do you need to know to determine the equation of a line?
- d. What information do you find in both (a) and (c)?

Answer.

- a. $h'(4) = \frac{3}{16}$.
- b. (i.) $P'(4) = 2(1.37)^4 \ln(1.37) \approx 2.218$ million cells per day; (ii.) the population is growing at an increasing rate.
- c. $y - 25 = -33(a + 1)$.
- d. The slope is a number, while the equation is, well, an equation.

Solution.

- a. Note that since $h(z) = z^{1/2} + z^{-1}$, we have $h'(z) = \frac{1}{2}z^{-1/2} - z^{-2}$. Thus, $h'(4) = \frac{1}{2(4)^{1/2}} - \frac{1}{4^2} = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$. Thus, the slope of the tangent line to $h(z)$ at the point where $z = 4$ is $\frac{3}{16}$.
- b. For (i.), note that $P'(t) = 2(1.37)^t \ln(1.37)$, and therefore $P'(4) = 2(1.37)^4 \ln(1.37) \approx 2.218$ million cells per day. For (ii.), we can compute $P''(t)$ and find that $P''(t) = 2(1.37)^t \ln(1.37) \ln(1.37)$ and therefore find that $P''(4) \approx 0.69825$. Since $P''(4)$ is positive, this tells us that the instantaneous rate of change $P'(t)$ is increasing at $t = 4$, and thus the population is growing at an increasing rate.
- c. Given $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$, first observe that $p(-1) = 3 + 2 + 7 + 1 + 12 = 25$, so the tangent line will pass through $(-1, 25)$. Further, since $p'(a) = 12a^3 - 6a^2 + 14a - 1$, we have $p'(-1) = -12 - 6 - 14 - 1 = -33$, which is the slope of the tangent line. The equation of the tangent line is therefore $y - 25 = -33(a + 1)$.
- d. The biggest difference is that (a) asks for the *slope* of the tangent line, while (c) asks for the *equation* of the tangent line. The latter requires the slope of the tangent line, but the slope and equation are two different entities.

2.1.4 Summary

- 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

$$\frac{d}{dx}[af(x) + bg(x)] = af'(x) + bg'(x).$$

2.1.5 Exercises

- (Derivative of a power function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Derivative of a rational function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Derivative of a root function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Derivative of a quadratic) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Derivative of a sum of power functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Simplifying a product before differentiating) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Simplifying a quotient before differentiating) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Finding a tangent line equation) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Determining where $f'(x) = 0$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- Let f and g be differentiable functions for which the following information is known: $f(2) = 5$, $g(2) = -3$, $f'(2) = -1/2$, $g'(2) = 2$.

- Let h be the new function defined by the rule $h(x) = 3f(x) - 4g(x)$. Determine $h(2)$ and $h'(2)$.
- Find an equation for the tangent line to $y = h(x)$ at the point $(2, h(2))$.
- 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?
- Estimate the value of $p(2.03)$ by using the local linearization of p at the point $(2, p(2))$.

Answer.

- $h(2) = 27; h'(2) = -19/2$.
- $L(x) = 27 - \frac{19}{2}(x - 2)$.
- p is increasing at $x = 2$.
- $p(2.03) \approx -11.56$.

Solution.

- First, since $h(x) = 3f(x) - 4g(x)$, we know $h(2) = 3f(2) - 4g(2) = 3(5) - 4(-3) = 27$. Next, by the sum and constant multiple rules, $h'(x) = 3f'(x) - 4g'(x)$. Hence, using the given derivative values, $h'(2) = 3f'(2) - 4g'(2) = 3(-1/2) - 4(2) = -19/2$.
- The tangent line at the point $(2, h(2))$ has equation $L(x) = h(2) + h'(2)(x - 2)$. Using the information determined in (a), we see $L(x) = 27 - \frac{19}{2}(x - 2)$.
- Since $p(x) = -2f(x) + \frac{1}{2}g(x)$, it follows that $p'(x) = -2f'(x) + \frac{1}{2}g'(x)$ by the sum and constant multiple rules. Using the given function and derivative values for f and g at $x = 2$, we see that $p(2) = -2f(2) + \frac{1}{2}g(2) = -2(5) + \frac{1}{2}(-3) = -\frac{23}{2}$ and $p'(2) = -2(-\frac{1}{2}) + \frac{1}{2}(2) = 2$. Since $p'(2) = 2 > 0$, we conclude that p is increasing at $x = 2$.
- First, the linearization of p at $a = 2$ is $L(x) = p(2) + p'(2)(x - 2) = -\frac{23}{2} + 2(x - 2)$. Using the linearization to estimate $p(2.03)$, we find that

$$p(2.03) \approx L(2.03) = -\frac{23}{2} + 2(2.03 - 2) = -11.5 + 2(0.03) = -11.56.$$

- Let functions p and q be the piecewise linear functions given by their respective graphs in Figure 2.1.10. Use the graphs to answer the following questions.

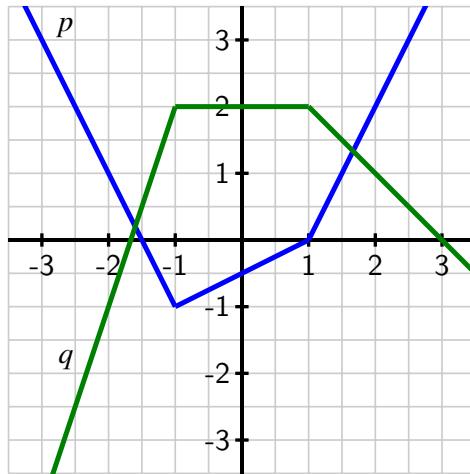


Figure 2.1.10: The graphs of p (in blue) and q (in green).

- At what values of x is p not differentiable? At what values of x is q not differentiable? Why?
- Let $r(x) = p(x) + 2q(x)$. At what values of x is r not differentiable? Why?
- Determine $r'(-2)$ and $r'(0)$.
- Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$.

Answer.

- p is not differentiable at $x = -1$ and $x = 1$; q is not differentiable at $x = -1$ and $x = 1$.
- r is not differentiable at $x = -1$ and $x = 1$.
- $r'(-2) = 4$; $r'(0) = 1$.
- $y = 4$.

Solution.

- The function p is differentiable at points where its graph is locally linear or smooth. Since there are sharp corners in the graph of p at $x = -1$ and $x = 1$, we see that p is not differentiable at those points. Similarly, the sharp corners in q at $x = -1$ and $x = 1$ mean that q is not differentiable at these points, either.
- In order for r to be differentiable at a point, both p and q must be differentiable at the point. So r is not differentiable at $x = -1$ and $x = 1$.
- Using the sum and scalar multiple rules, we have $r'(x) = p'(x) + 2q'(x)$. The slope of p at $x = -2$ is -2 and the slope of q at $x = -2$ is 3 , so $p'(-2) = -2$ and $q'(-2) = 3$. It follows that $r'(-2) = p'(-2) + 2q'(-2) = (-2) + 2(3) = 4$. Similarly, the slope of p at $x = 0$ is 1 and the slope of q at $x = 0$ is 0 , so $p'(0) = 1$ and $q'(0) = 0$. It follows that $r'(0) = p'(0) + 2q'(0) = (1) + 2(0) = 1$.
- Note that $r(2) = p(2) + 2q(2) = 2 + 2(1) = 4$ and $r'(2) = p'(2) + 2q'(2) = 2 + 2(-1) = 0$. Thus, the line tangent to the graph of r at $x = 2$ is $y = r(2) + r'(2)(x - 2) = 4$.

12. 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 $0 < t < 1$.

- Let $w(t) = 3t^t - 2\arccos(t)$. Determine $w'(t)$.
- Find an equation for the tangent line to $y = w(t)$ at the point $(\frac{1}{2}, w(\frac{1}{2}))$.
- Let $v(t) = t^t + \arccos(t)$. Is v increasing or decreasing at the instant $t = \frac{1}{2}$? Why?

Answer.

- $w'(t) = 3t^t(\ln(t) + 1) + 2\frac{1}{\sqrt{1-t^2}}$.
- $L(t) = (\frac{3}{\sqrt{2}} - \frac{2\pi}{3}) + (\frac{3}{\sqrt{2}}(\ln(\frac{1}{2}) + 1) + \frac{4}{\sqrt{3}})(t - 2)$.
- v is increasing at $t = \frac{1}{2}$.

Solution.

- We note that w has the structure $w(t) = 3r(t) - 2s(t)$, so by the sum and constant multiple rules, $w'(t) = 3r'(t) - 2s'(t)$. Applying the given information about r' and s' , $w'(t) = 3t^t(\ln(t) + 1) + 2\frac{1}{\sqrt{1-t^2}}$.
- We see that $w(\frac{1}{2}) = 3(\frac{1}{2})^{\frac{1}{2}} - 2\arccos(\frac{1}{2}) = 3 \cdot \frac{1}{\sqrt{2}} - 2 \cdot \frac{\pi}{3} = \frac{3}{\sqrt{2}} - \frac{2\pi}{3}$. Similarly,

$$\begin{aligned} w'(\frac{1}{2}) &= 3\left(\frac{1}{2}\right)^{\frac{1}{2}}\left(\ln\left(\frac{1}{2}\right) + 1\right) + 2\frac{1}{\sqrt{1-(\frac{1}{2})^2}} \\ &= \frac{3}{\sqrt{2}}\left(\ln\left(\frac{1}{2}\right) + 1\right) + \frac{4}{\sqrt{3}}. \end{aligned}$$

Using these values, we find that the tangent line to w at the point $(\frac{1}{2}, w(\frac{1}{2}))$ is

$$L(t) = w\left(\frac{1}{2}\right) + w'\left(\frac{1}{2}\right)\left(t - \frac{1}{2}\right) = \left(\frac{3}{\sqrt{2}} - \frac{2\pi}{3}\right) + \left(\frac{3}{\sqrt{2}}\left(\ln\left(\frac{1}{2}\right) + 1\right) + \frac{4}{\sqrt{3}}\right)(t - 2).$$

- To determine whether v is increasing or decreasing at $t = \frac{1}{2}$, we compute $v'(\frac{1}{2})$. First, $v'(t) = t^t(\ln(t) + 1) + \frac{1}{\sqrt{1-t^2}}$. Hence,

$$v'\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{\frac{1}{2}}\left(\ln\left(\frac{1}{2}\right) + 1\right) + \frac{1}{\sqrt{1-(\frac{1}{2})^2}}.$$

Since $(\ln(\frac{1}{2}) + 1)$ is positive, it's clear that $v'(\frac{1}{2}) > 0$, and thus v is increasing at $t = \frac{1}{2}$.

13. 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.

- Use the limit definition of the derivative to show that

$$f'(x) = \lim_{h \rightarrow 0} \frac{a^x \cdot a^h - a^x}{h}.$$

- b. Explain why it is also true that

$$f'(x) = a^x \cdot \lim_{h \rightarrow 0} \frac{a^h - 1}{h}.$$

- c. Use computing technology and small values of h to estimate the value of

$$L = \lim_{h \rightarrow 0} \frac{a^h - 1}{h}$$

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

$$L = \lim_{h \rightarrow 0} \frac{a^h - 1}{h} = 1.$$

- 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 $\frac{d}{dx}[3^x]$?

- f. How do your investigations in (d) lead to a particularly important fact about the function $f(x) = e^x$?

Answer.

- a.

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{a^{x+h} - a^x}{h} \\ &= \lim_{h \rightarrow 0} \frac{a^x \cdot a^h - a^x}{h} \\ &= \lim_{h \rightarrow 0} \frac{a^x(a^h - 1)}{h}, \end{aligned}$$

- b. Since a^x does not depend at all on h , we may treat a^x as constant in the noted limit and thus write the value a^x in front of the limit being taken.

- c. When $a = 2$, $L \approx 0.6931$; when $a = 3$, $L \approx 1.0986$.

- d. $a \approx 2.71828$ (for which $L \approx 1.000$)

- e. $\frac{d}{dx}[2^x] = 2^x \cdot \ln(2)$ and $\frac{d}{dx}[3^x] = 3^x \cdot \ln(3)$

- f.

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

Solution.

- a. By definition,

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

Using the fact that $f(x) = a^x$ along with $a^{x+h} = a^x \cdot a^h$, we have

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{a^{x+h} - a^x}{h} \\ &= \lim_{h \rightarrow 0} \frac{a^x \cdot a^h - a^x}{h} \\ &= \lim_{h \rightarrow 0} \frac{a^x(a^h - 1)}{h}, \end{aligned}$$

where the last equality follows by removing the common factor of a^x from each term in the numerator.

- b. Since a^x does not depend at all on h , we may treat a^x as constant in the noted limit. Thus,

$$f'(x) = \lim_{h \rightarrow 0} \frac{a^x(a^h - 1)}{h} = a^x \cdot \lim_{h \rightarrow 0} \frac{a^h - 1}{h}.$$

Note that this tells us that the derivative of a^x is some constant multiple of a^x (and that the constant's values must be $\lim_{h \rightarrow 0} \frac{a^h - 1}{h}$.)

- c. Using $h = \pm 0.0001$, it appears that when $a = 2$, $L \approx 0.6931$, while when $a = 3$, $L \approx 1.0986$.
- d. By experimenting, we see first that $a \approx 2.7$ (for which $L \approx 0.993$, so a needs to be slightly bigger than 2.7). Continuing to try values of a that result in L being closer and closer to 1, we find $a \approx 2.71828$ (for which $L \approx 1.000$)
- e. We note that $\ln(2) \approx 0.6931$ and $\ln(3) \approx 1.0986$. Since these values correspond to the values of

$$\lim_{h \rightarrow 0} \frac{a^h - 1}{h}$$

when $a = 2$ and $a = 3$, we conjecture that

$$\frac{d}{dx}[2^x] = 2^x \cdot \ln(2)$$

and

$$\frac{d}{dx}[3^x] = 3^x \cdot \ln(3)$$

- f. Our work suggests that $\frac{d}{dx}[a^x] = a^x \cdot \ln(a)$, and when $a = e \approx 2.71828$, it follows that

$$\frac{d}{dx}[e^x] = e^x \cdot \ln(e) = e^x$$

Thus, e^x is a very special function: its derivative is itself!

2.2 The product and quotient rules

Motivating Questions

- How does the algebraic structure of a function guide 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 differentiate quickly?
-

So far, we can differentiate 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.

Example 2.2.1. Differentiate

$$f(x) = 7x^{11} - 4 \cdot 9^x + \pi \sin(x) - \sqrt{3} \cos(x)$$

Because f is a sum of basic functions, we can now quickly say that $f'(x) = 77x^{10} - 4 \cdot 9^x \ln(9) + \pi \cos(x) + \sqrt{3} \sin(x)$.

What about a product or quotient of two basic functions, such as

$$p(z) = z^3 \cos(z),$$

or

$$q(t) = \frac{\sin(t)}{2^t}?$$

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.

Example 2.2.2. Let f and g be the functions defined by $f(t) = 2t^2$ and $g(t) = t^3 + 4t$.

- Determine $f'(t)$ and $g'(t)$.
- 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.
- True or false: $p'(t) = f'(t) \cdot g'(t)$.

- d. Let $q(t) = \frac{t^3+4t}{2t^2}$ and observe that $q(t) = \frac{g(t)}{f(t)}$. 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: $q'(t) = \frac{g'(t)}{f'(t)}$.

2.2.1 The product rule

As part (b) of Example 2.2.2 shows, it is not true in general that the derivative of a product of two functions is the product of the derivatives of those functions. To see why this is the case, we consider an example involving meaningful functions.

Say that an investor is regularly purchasing stock in a particular company. Let $N(t)$ represent the number of shares owned on day t , where $t = 0$ represents the first day on which shares were purchased. Let $S(t)$ give the value of one share of the stock on day t ; note that the units on $S(t)$ are dollars per share. To compute the total value of the stock on day t , we take the product

$$V(t) = N(t) \text{ shares} \cdot S(t) \text{ dollars per share},$$

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 is changing, while $S'(t)$ measures the rate at which the value per share is changing. How do these respective rates of change affect the rate of change of the total value function?

To help us 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 $S(100) = 27.50$.
- 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).
- On that same day the price of the stock is rising at a rate of 0.75 dollars per share per day.

In calculus notation, the latter two facts tell 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 increasing (and the value of each share is constant), the rate at which total value would rise is the product of the current value of the shares and 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 show 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, the rate at which the total value would rise is the product of the number of shares and the rate of change of share value. The total value is rising at a rate of

$$N(100) \cdot S'(100) = 520 \text{ shares} \cdot 0.75 \frac{\text{dollars per share}}{\text{day}} = 390 \frac{\text{dollars}}{\text{day}}.$$

Of course, when both the number of shares and the value of each share are changing, we have to include both of these sources. In that case the rate at which the total value is rising is

$$V'(100) = S(100) \cdot N'(100) + N(100) \cdot S'(100) = 330 + 390 = 720 \frac{\text{dollars}}{\text{day}}.$$

We expect the total value of the investor's holdings to rise by about \$720 on the 100th day.¹

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 $P(x) = f(x) \cdot g(x)$, 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.

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

$$P'(x) = f(x)g'(x) + g(x)f'(x).$$

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. In words 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 2.2.3. If $P(z) = z^3 \cdot \cos(z)$, we can use the product rule to differentiate P . 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,

$$P'(z) = z^3(-\sin(z)) + \cos(z)3z^2 = -z^3 \sin(z) + 3z^2 \cos(z).$$

¹While 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 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.

Example 2.2.4. Use the product rule to answer each of the questions below. Throughout, be sure to carefully label any derivative you find by name. It is not necessary to algebraically simplify any of the derivatives you compute.

- Let $m(w) = 3w^{17}4^w$. Find $m'(w)$.
- Let $h(t) = (\sin(t) + \cos(t))t^4$. Find $h'(t)$.
- 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 $f(x) = e^x \sin(x)$.
- 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 $g(x) = (x^2 + x)2^x$.

Hint.

- Let the first function be $3w^{17}$.
- Let the first function be $(\sin(t) + \cos(t))$.
- Remember that the slope of the tangent line to $y = f(x)$ at $(a, f(a))$ is given by $f'(a)$.
- Remember that $L(x) = g(a) + g'(a)(x - a)$.

Answer.

- $m'(w) = 3w^{17} \cdot 4^w \ln(4) + 4^w \cdot 51w^{16}$.
- $h'(t) = (\sin(t) + \cos(t)) \cdot 4t^3 + t^4 \cdot (\cos(t) - \sin(t))$.
- $f'(1) = e(\cos(1) + \sin(1)) \approx 3.756$.
- $L(x) = -\frac{1}{2}(x + 1)$.

Solution.

- By the product rule,

$$m'(w) = 3w^{17} \cdot 4^w \ln(4) + 4^w \cdot 51w^{16}.$$

- By the product rule,

$$h'(t) = (\sin(t) + \cos(t)) \cdot 4t^3 + t^4 \cdot (\cos(t) - \sin(t)).$$

- To determine the slope of the tangent line at $a = 1$, we want to find $f'(1)$. Since $f(x) = e^x \sin(x)$, the product rule tells us that $f'(x) = e^x \cdot \cos(x) + \sin(x) \cdot e^x$. Thus, $f'(1) = e^1 \cdot \cos(1) + \sin(1) \cdot e^1 = e(\cos(1) + \sin(1)) \approx 3.756$.
- First, observe that $g(-1) = ((-1)^2 - 1) \cdot 2^{-1} = 0$. Further, by the product rule, $g'(x) = (x^2 + x) \cdot 2^x \ln(2) + 2^x \cdot (2x + 1)$, and therefore $g'(-1) = ((-1)^2 - 1) \cdot 2^{-1} \ln(2) + 2^{-1} \cdot (2(-1) + 1) = 0 + \frac{1}{2}(-1) = -\frac{1}{2}$. Therefore,

$$L(x) = g(-1) + g'(-1)(x + 1) = 0 - \frac{1}{2}(x + 1) = -\frac{1}{2}(x + 1).$$

2.2.2 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. Let $Q(x)$ be defined by $Q(x) = f(x)/g(x)$, where f and g are both differentiable functions. It turns out that Q is differentiable everywhere that $g(x) \neq 0$. We would like a formula for Q' in terms of f , g , f' , and g' . multiplying both sides of the formula $Q = f/g$ by g , we observe that

$$f(x) = Q(x) \cdot g(x).$$

Now we can use the product rule to differentiate f .

$$f'(x) = Q(x)g'(x) + g(x)Q'(x).$$

We want to know a formula for Q' , so we solve this equation for $Q'(x)$.

$$Q'(x)g(x) = f'(x) - Q(x)g'(x)$$

and dividing both sides by $g(x)$, we have

$$Q'(x) = \frac{f'(x) - Q(x)g'(x)}{g(x)}.$$

Finally, we recall that $Q(x) = \frac{f(x)}{g(x)}$. Substituting this expression in the preceding equation, 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 calculation gives us the *quotient rule*.

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

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

As with 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 2.2.5. If $Q(t) = \sin(t)/2^t$, we call $\sin(t)$ the top function and 2^t the bottom function. By the quotient rule, Q' is 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, so 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 execute the quotient rule correctly but then make an algebra error.

Example 2.2.6. 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.

- Let $r(z) = \frac{3^z}{z^4+1}$. Find $r'(z)$.
- Let $v(t) = \frac{\sin(t)}{\cos(t)+t^2}$. Find $v'(t)$.
- Determine the slope of the tangent line to the curve $R(x) = \frac{x^2 - 2x - 8}{x^2 - 9}$ at the point where $x = 0$.
- 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.

Hint.

- When applying the quotient rule, use parentheses around the bottom function, $z^4 + 1$, to ensure that when you compute "the bottom times the derivative of the top" that the rule is applied correctly.
- When applying the quotient rule, use parentheses around the bottom function, $\cos(t) + t^2$, and its derivative to ensure that the rule is applied correctly.
- Remember one of the key interpretations of the derivative.
- Let the top function be $100t$ and simply use the constant multiple rule to find its derivative.

Answer.

- a. $r'(z) = \frac{(z^4+1)3^z \ln(3) - 3^z(4z^3)}{(z^4+1)^2}$.
- b. $v'(t) = \frac{(\cos(t)+t^2)\cos(t) - \sin(t)(\sin(t)+2t)}{(\cos(t)+t^2)^2}$.
- c. $R'(0) = \frac{2}{9}$.
- d. $I'(0.5) = \frac{50}{e^{0.5}} \approx 30.327$, $I'(2) = \frac{-100}{e^2} \approx -13.534$, and $I'(5) = \frac{-400}{e^4} \approx -2.695$, each in candles per millisecond.

Solution.

- a. By the quotient rule,

$$r'(z) = \frac{(z^4+1)3^z \ln(3) - 3^z(4z^3)}{(z^4+1)^2}.$$

- b. By the quotient rule,

$$v'(t) = \frac{(\cos(t)+t^2)\cos(t) - \sin(t)(\sin(t)+2t)}{(\cos(t)+t^2)^2}.$$

- c. We first compute $R'(x)$. By the quotient rule,

$$R'(x) = \frac{(x^2-9)(2x-2) - (x^2-2x-8)(2x)}{(x^2-9)^2}.$$

From this, it follows that $R'(0) = \frac{(-9)(-2)-(-8)(0)}{(-9)^2} = \frac{2}{9}$, which is the slope of the tangent line to the curve at the point where $x = 0$.

- d. By the quotient rule and algebraic simplification,

$$I'(t) = \frac{e^t 100 - 100t e^t}{(e^t)^2} = \frac{100 - 100t}{e^t}.$$

Thus, $I'(0.5) = \frac{50}{e^{0.5}} \approx 30.327$, $I'(2) = \frac{-100}{e^2} \approx -13.534$, and $I'(5) = \frac{-400}{e^4} \approx -2.695$, each measured in candles per millisecond. These results show that at $t = 0.5$, the intensity of the flash is increasing rapidly, while at $t = 2$ and $t = 5$, the intensity is decreasing, with the intensity decreasing more rapidly when $t = 2$.

2.2.3 Combining rules

In order to apply the derivative shortcut rules correctly we must recognize the fundamental structure of a function.

Example 2.2.7. Determine the derivative of the function

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

How do we decide which rules to apply? Our first task is to recognize the structure of the function. This function f is a sum of two slightly less complicated functions, so we can apply the sum rule² to get

$$\begin{aligned} f'(x) &= \frac{d}{dx} \left[x \sin(x) + \frac{x^2}{\cos(x) + 2} \right] \\ &= \frac{d}{dx} [x \sin(x)] + \frac{d}{dx} \left[\frac{x^2}{\cos(x) + 2} \right] \end{aligned}$$

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

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

Example 2.2.8. Differentiate

$$s(y) = \frac{y \cdot 7^y}{y^2 + 1}.$$

The function s is a quotient of two simpler functions, so the quotient rule will be needed. To begin, we set up the quotient rule and use the notation $\frac{d}{dy}$ to indicate the derivatives of the numerator and denominator. Thus,

$$s'(y) = \frac{(y^2 + 1) \cdot \frac{d}{dy} [y \cdot 7^y] - y \cdot 7^y \cdot \frac{d}{dy} [y^2 + 1]}{(y^2 + 1)^2}.$$

Now, there remain two derivatives to calculate. The first one, $\frac{d}{dy} [y \cdot 7^y]$ calls for use of the product rule, while the second, $\frac{d}{dy} [y^2 + 1]$ needs only 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 simplification is possible, we are content to leave $s'(y)$ in its current form.

Success in applying derivative rules begins with recognizing the structure of the function, followed by the careful and diligent application of the relevant derivative rules. The best way to become proficient at this process is to do a large number of examples.

Example 2.2.9. 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 $f'(r)$.

²When 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.

- b. Let $p(t) = \frac{\cos(t)}{t^6 \cdot 6^t}$. Find $p'(t)$.
- c. Let $g(z) = 3z^7 e^z - 2z^2 \sin(z) + \frac{z}{z^2+1}$. Find $g'(z)$.
- d. A moving particle has its position in feet at time t in seconds given by the function $s(t) = \frac{3\cos(t)-\sin(t)}{e^t}$. 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 $q(x) = \frac{f(x)}{g(x)}$, calculate $p'(3)$ and $q'(3)$.

Hint.

- a. Observe that f is fundamentally a product. Which is the first function? The second?
- b. Note that p has the overall structure of a quotient.
- c. Think about how g is a sum of three functions. What is the structure of each of the three functions in the sum?
- d. How is the velocity of a moving object related to its position?
- e. Since we know $p(x) = f(x) \cdot g(x)$, it follows $p'(x) = f(x)g'(x) + g(x)f'(x)$.

Answer.

- a. $f'(r) = (5r^3 + \sin(r))[4^r \ln(4) + 2 \sin(r)] + (4^r - 2 \cos(r))[15r^2 + \cos(r)]$.
- b. $p'(t) = \frac{t^6 \cdot 6^t[-\sin(t)] - \cos(t)[t^6 \cdot 6^t \ln(6) + 6^t \cdot 6t^5]}{(t^6 \cdot 6^t)^2}$.
- c. $g'(z) = 3[z^7 e^z + 7z^6 e^z] - 2[z^2 \cos(z) + 2z \sin(z)] + \frac{(z^2+1)1-z(2z)}{(z^2+1)^2}$.
- d. $s'(1) = \frac{-2\sin(1)-4\cos(1)}{e^1} \approx -1.414$ feet per second.
- e. $p'(3) = 30$ and $q'(3) = \frac{13}{8}$.

Solution.

- a. Using the product rule, followed by the sum and constant multiple rule, observe that

$$\begin{aligned} f'(r) &= (5r^3 + \sin(r)) \frac{d}{dr}[4^r - 2 \cos(r)] + (4^r - 2 \cos(r)) \frac{d}{dr}[5r^3 + \sin(r)] \\ &= (5r^3 + \sin(r))[4^r \ln(4) + 2 \sin(r)] + (4^r - 2 \cos(r))[15r^2 + \cos(r)] \end{aligned}$$

- b. We use the quotient rule on p , followed by the product rule to differentiate the denominator, finding that

$$\begin{aligned} p'(t) &= \frac{t^6 \cdot 6^t \frac{d}{dt}[\cos(t)] - \cos(t) \frac{d}{dt}[t^6 \cdot 6^t]}{(t^6 \cdot 6^t)^2} \\ &= \frac{t^6 \cdot 6^t[-\sin(t)] - \cos(t)[t^6 \cdot 6^t \ln(6) + 6^t \cdot 6t^5]}{(t^6 \cdot 6^t)^2} \end{aligned}$$

c. Using the sum and constant multiple rules, it follows first that

$$g'(z) = 3 \frac{d}{dz}[z^7 e^z] - 2 \frac{d}{dz}[z^2 \sin(z)] + \frac{d}{dz} \left[\frac{z}{z^2 + 1} \right].$$

Applying the product rule in the first two terms and the quotient rule in the third, we find that

$$g'(z) = 3[z^7 e^z + 7z^6 e^z] - 2[z^2 \cos(z) + 2z \sin(z)] + \frac{(z^2 + 1)1 - z(2z)}{(z^2 + 1)^2}.$$

d. The particle's instantaneous velocity at the moment $t = 1$ is given by $s'(1)$. We use the quotient rule to find $s'(t)$, and simplify by removing a common factor of e^t to get

$$s'(t) = \frac{e^t(-3 \sin(t) - \cos(t)) - (3 \cos(t) - \sin(t))e^t}{(e^t)^2} = \frac{-2 \sin(t) - 4 \cos(t)}{e^t}.$$

Thus, $s'(1) = \frac{-2 \sin(1) - 4 \cos(1)}{e^1} \approx -1.414$, which is the particle's instantaneous velocity in feet per second at the moment $t = 1$.

e. Since $p(x) = f(x) \cdot g(x)$, the product rule tells us

$$p'(x) = f(x)g'(x) + g(x)f'(x),$$

and since $q(x) = \frac{f(x)}{g(x)}$, by the quotient rule we know

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

Using the given information ($f(3) = -2$, $f'(3) = 7$, $g(3) = 4$, and $g'(3) = -1$) we now see that

$$p'(3) = f(3)g'(3) + g(3)f'(3) = (-2)(-1) + (4)(7) = 30$$

and

$$q'(3) = \frac{g(3)f'(3) - f(3)g'(3)}{g(3)^2} = \frac{(4)(7) - (-2)(-1)}{4^2} = \frac{13}{8}.$$

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 $(a, f(a))$.

2.2.4 Summary

- If a function is a sum, product, or quotient of simpler functions, then we can use the sum, product, or quotient rules to differentiate it 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

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

- Along with the constant multiple and sum rules, the product and quotient rules enable us to compute the derivative of any function that consists of sums, constant multiples, products, and quotients of basic functions. For instance, if F has the form

$$F(x) = \frac{2a(x) - 5b(x)}{c(x) \cdot d(x)},$$

then F is a quotient, in which the numerator is a sum of constant multiples and the denominator is a product. This, 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.

2.2.5 Exercises

1. (Derivative of a basic product) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Derivative of a product) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Derivative of a quotient of linear functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Derivative of a rational function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Derivative of a product of trigonometric functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Derivative of a product of power and trigonometric functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Derivative of a sum that involves a product) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. (Product and quotient rules with graphs) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
9. (Product and quotient rules with given function values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
10. Let f and g be differentiable functions for which the following information is known: $f(2) = 5$, $g(2) = -3$, $f'(2) = -1/2$, $g'(2) = 2$.

- Let h be the new function defined by the rule $h(x) = g(x) \cdot f(x)$. Determine $h(2)$ and $h'(2)$.
- Find an equation for the tangent line to $y = h(x)$ at the point $(2, h(2))$ (where h is the function defined in (a)).
- 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?
- Estimate the value of $r(2.06)$ (where r is the function defined in (c)) by using the local linearization of r at the point $(2, r(2))$.

Answer.

- $h(2) = -15; h'(2) = 23/2$.
- $L(x) = -15 + 23/2(x - 2)$.
- Increasing.
- $r(2.06) \approx -0.5796$.

Solution.

- Since $h(x) = g(x) \cdot f(x)$, it follows $h(2) = g(2) \cdot f(2) = (5) \cdot (-3) = -15$. Further, by the product rule, we know that $h'(x) = g(x) \cdot f'(x) + f(x) \cdot g'(x)$. Using the given function and derivative values, $h'(2) = g(2) \cdot f'(2) + f(2) \cdot g'(2) = (-3)(-1/2) + (5)(2) = 23/2$.
- The tangent line to h at $a = 2$ is given by $L(x) = h(2) + h'(2)(x - 2)$, so using our results from (a), we have $L(x) = -15 + 23/2(x - 2)$.
- Since $r(x) = \frac{g(x)}{f(x)}$, but the quotient rule that $r'(x) = \frac{f(x) \cdot g'(x) - g(x) \cdot f'(x)}{(f(x))^2}$, and thus $r'(2) = \frac{f(2) \cdot g'(2) - g(2) \cdot f'(2)}{(f(2))^2}$. Employing the known function and derivative values, $r'(2) = \frac{(5)(2) - (-3)(-1/2)}{(5)^2} = \frac{17/2}{25} = \frac{17}{50}$. Therefore, $r'(2) > 0$, and r is increasing at $a = 2$.
- The local linearization of r at $a = 2$ is $L(x) = r(2) + r'(2)(x - 2) = -\frac{3}{5} + \frac{17}{20}(x - 2)$, and thus $r(2.06) \approx L(2.06) = -\frac{3}{5} + \frac{17}{50}(2.06 - 2) = -0.5796$.

11. 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 $0 < t < 1$.

- Let $w(t) = t^t \arccos(t)$. Determine $w'(t)$.
- Find an equation for the tangent line to $y = w(t)$ at the point $(\frac{1}{2}, w(\frac{1}{2}))$.
- Let $v(t) = \frac{t^t}{\arccos(t)}$. Is v increasing or decreasing at the instant $t = \frac{1}{2}$? Why?

Answer.

- $w'(t) = t^t(\ln t + 1) \cdot (\arccos t) + t^t \cdot \frac{-1}{\sqrt{1-t^2}}$.
- $L(t) \approx 0.740 - 0.589(t - 0.5)$.
- Increasing.

Solution.

- a. Using the Product Rule and the information given,

$$\begin{aligned} w'(t) &= \frac{d}{dt} (t^t) \cdot (\arccos t) + t^t \cdot \frac{d}{dt} (\arccos t) \\ &= t^t (\ln t + 1) \cdot (\arccos t) + t^t \cdot \frac{-1}{\sqrt{1-t^2}} \end{aligned}$$

- b. Using the local linearization formula, we have $L(t) = w\left(\frac{1}{2}\right) + w'\left(\frac{1}{2}\right)(t - \frac{1}{2})$. Substituting $t = \frac{1}{2}$ into the formulas for w and w' and using a computational device gives

$$w(2) = \left(\frac{1}{2}\right)^{\left(\frac{1}{2}\right)} \arccos\left(\frac{1}{2}\right) \approx 0.740$$

and

$$w'(2) = \left[\left(\frac{1}{2}\right)^{\left(\frac{1}{2}\right)} (\ln t + 1) \cdot \left(\arccos\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2}\right)^{\left(\frac{1}{2}\right)} \cdot \frac{-1}{\sqrt{1-\left(\frac{1}{2}\right)^2}} \right] \approx -0.589.$$

Thus, $L(t) \approx 0.740 - 0.589(t - 0.5)$.

- c. Using the Quotient Rule,

$$v'(t) = \frac{p(t)^t (\ln(t) + 1) (\arccos(t)) - (t)^t \cdot \left(\frac{-1}{\sqrt{1-(t)^2}}\right) \arccos^2(t)}{\sqrt{1-(t)^2}}$$

Using a computational device to evaluate $v'(2)$, it follows $v'(2) \approx 0.952$. Since this number is positive, the instantaneous rate of change of v at $t = \frac{1}{2}$ is positive, which means that v is increasing at that instant.

12. Let functions p and q be the piecewise linear functions given by their respective graphs in Figure 2.2.10. Use the graphs to answer the following questions.

- Let $r(x) = p(x) \cdot q(x)$. Determine $r'(-2)$ and $r'(0)$.
- Are there values of x for which $r'(x)$ does not exist? If so, which values, and why?
- Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$.
- Let $z(x) = \frac{q(x)}{p(x)}$. Determine $z'(0)$ and $z'(2)$.
- Are there values of x for which $z'(x)$ does not exist? If so, which values, and why?

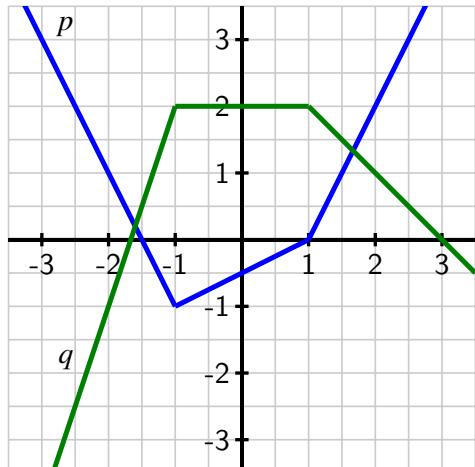


Figure 2.2.10: The graphs of p (in blue) and q (in green).

Answer.

- a. $r'(-2) = 2$ and $r'(0) = 0.25$.
- b. At $x = -1$ and $x = 1$.
- c. $L(x) = 2$.
- d. $z'(0) = \frac{1}{16}$ and $z'(2) = -1$.
- e. At $x = -1, x = 1, x = -1.5$, and $x = 1$.

Solution.

- a. For all values of t for which , the product rule applies, and we have $r'(x) = p(x) \cdot q'(x) + q(x) \cdot p'(x)$. Since p and q are differentiable at both $x = -2$ and $x = 0$, we can determine the values of p' and q' at these points by computing the slope of the function (which is piecewise linear). In particular, $p'(-2) = -2$, $q'(-2) = 3$, $p'(0) = 0$, and $q'(0) = \frac{1}{2}$. In addition, we can read the values of p and q at these points from the graph: $p(-2) = 1$, $q(-2) = -1$, $p(0) = \frac{1}{2}$, and $q(0) = 2$. Thus,

$$\begin{aligned} r'(-2) &= p(-2) \cdot q'(-2) + q(-2) \cdot p'(-2) \\ &= (1)(3) + (-1)(-2) \\ &= 2 \end{aligned}$$

and

$$\begin{aligned} r'(0) &= p(0) \cdot q'(0) + q(0) \cdot p'(0) \\ &= (0.5)(0.5) + (2)(0) \\ &= 0.25. \end{aligned}$$

- b. Neither $r'(-1)$ nor $r'(1)$ exist because both p and q fail to be differentiable at $x = -1$ and $x = 1$ due to the sharp corners on their graphs.
- c. To find the tangent line to r at $(2, r(2))$, we first observe that $r(2) = p(2) \cdot q(2) = 2 \cdot 1 = 2$. In addition, we know by the product rule that $r'(2) = p(2) \cdot q'(2) + q(2) \cdot p'(2) = (2)(-1) + (1)(2) = 0$. Thus, the tangent line is given by $L(x) = r(2) + r'(2)(x - 2) = 2 + 0(x - 2) = 2$. In particular, the tangent line is horizontal at $(2, r(2))$.
- d. By the quotient rule,

$$z'(x) = \frac{p(x) \cdot q'(x) - q(x) \cdot p'(x)}{p(x)^2}$$

Thus, using the values of p and q and their derivatives at $x = 0$ and $x = 2$ that we have determined in earlier parts of this problem,

$$z'(0) = \frac{p(0) \cdot q'(0) - q(0) \cdot p'(0)}{p(0)^2} = \frac{(0.5)(0.5) - (2)(0)}{2^2} = \frac{1}{16}$$

and

$$z'(2) = \frac{p(2) \cdot q'(2) - q(2) \cdot p'(2)}{p(2)^2} = \frac{(2)(-1) - (1)(2)}{2^2} = -1.$$

- e. z fails to be differentiable at $x = -1$ and $x = 1$ because both p and q have sharp corners at these x -values. In addition, since $p(-1.5) = 0$ and $p(1) = 0$, z is not differentiable at $x = -1.5$ and $x = 1$.

13. 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.

- 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?
- What is the value of $C(0)$? What does it measure?
- Write an expression for $C'(t)$ in terms of $A(t)$, $A'(t)$, $Y(t)$, and $Y'(t)$. Explain your thinking.
- What is the value of $C'(0)$? What does it measure?
- Based on the given information and your work above, estimate the value of $C(1)$.

Answer.

- $C(t) = A(t)Y(t)$ bushels in year t .
- 1190000 bushels of corn.
- $C'(t) = A(t)Y'(t) + A'(t)Y(t)$.
- $C'(0) = 158000$ bushels per year.
- $C(1) \approx 1348000$ bushels.

Solution.

- If we multiply $A(t)$ acres by $Y(t)$ bushels per acre, we arrive at $C(t) = A(t)Y(t)$ bushels as the total number of bushels of corn the farmer produces in year t .
- We are given that $A(0) = 7000$ acres with a yield of $Y(0) = 170$ bushels per acre, giving us $C(0) = A(0)Y(0) = (7000)(170) = 1190000$ bushels. Thus, the farmer produces 1190000 bushels of corn in the present year.
- The product rule tells us that $C'(t) = A(t)Y'(t) + A'(t)Y(t)$.
- We know $A(0) = 7000$ and $Y(0) = 170$. We are also given that the farmer will 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. This tells us that $A'(t) = 600$ and $Y'(0) = 8$. So

$$C'(0) = A(0)Y'(0) + A'(0)Y(0) = (7000)(8) + (600)(170) = 158000.$$

Note that the units of $A(t)$ are acres and the units of $Y'(t)$ are bushels per acre per year. So the units of $A(0)Y'(0)$ are bushels per year. Similarly, the units of $A'(t)$ are acres per year and the units of $Y(t)$ are bushels per acre, so the units of $A'(0)Y(0)$ are also bushels per acre. So the units of $C'(0)$ are bushels per acre. Therefore, $C'(0) = 158000$ tells us that the number of bushels the farmer can produce will increase by approximately 158000 bushels for each year past the present year.

- e. The linearization of C at $t = 0$ is $L(t) = C(0) + C'(0)(t - 0) = 1190000 + 158000t$, $C(1) \approx L(1) = 1190000 + 158000(1) = 1348000$ bushels.
- 14.** 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 $f'(80) = 0.0004$.
- 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 $g'(80)$.
 - 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 $h'(80)$.
 - 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.

Answer.

- $g(80) = 20$ kilometers per liter, and $g'(80) = -0.16$. kilometers per liter per kilometer per hour.
- $h(80) = 4$ liters per hour and $h'(80) = 0.082$ liters per hour per kilometer per hour.
- Think carefully about units and how each of the three pairs of values expresses fundamentally the same facts.

Solution.

- Observe that the units on $g(v)$ are “kilometers per liter.” Since the units on $f(v)$ are “liters per kilometer,” it follows that g is the reciprocal of f : $g(v) = \frac{1}{f(v)}$. By the quotient rule, we then know that

$$g'(v) = \frac{f(v) \cdot 0 - 1 \cdot f'(v)}{f(v)^2} = -\frac{f'(v)}{f(v)^2}.$$

Further, using the given values we can say that $g(80) = \frac{1}{f(80)} = \frac{1}{0.05} = 20$ kilometers per liter, and $g'(80) = -\frac{f'(80)}{f(80)^2} = -\frac{0.0004}{0.05^2} = -0.16$. Note that the units on $g'(80)$ are “kilometers per liter per kilometer per hour.”

- Since the units on $h(v)$ are “liters per hour,” and the units on $f(v)$ are “liters per kilometer,” we note that if we multiply $f(v)$ by v (whose units are “kilometers per hour”), we get that the units on $v \cdot f(v)$ match the units of $h(v)$. Thus,

$$h(v) = v \cdot f(v).$$

By the product rule, it follows that

$$h'(v) = v \cdot \frac{d}{dv}[f(v)] + \frac{d}{dv}[v] \cdot f(v) = v \cdot f'(v) + 1 \cdot f(v).$$

Using the known function and derivative values of f , we determine that $h(80) = 80 \cdot 0.05 = 4$ liters per hour and $h'(80) = 80 \cdot 0.0004 + 0.05 = 0.082$ liters per hour per kilometer per hour.

c. To explain the meaning of these function and derivative values to someone who doesn't know calculus, we could say something like the following:

- The given values of $f(80) = 0.05$ and $f'(80) = 0.0004$ tell us that when the car is going 80 kilometers per hour, it is using 0.05 liters of fuel for every kilometer it travels. In addition, for each additional kilometer per hour of velocity above 80 kph, we expect that the car will use about an additional 0.0004 liters of fuel per kilometer.
- The values of $g(80) = 20$ and $g'(80) = -0.16$ tell us that when the car is going 80 kilometers per hour, it is able to travel 20 kilometers for each liter of fuel used. In addition, for each additional kilometer per hour of velocity above 80 kph, we expect that the car will be able to travel about 0.16 fewer kilometers per liter of fuel.
- Finally, the values of $h(80) = 4$ and $h'(80) = 0.82$ tell us that when the car is going 80 kilometers per hour, it is using 4 liters of fuel for every hour it travels. In addition, for each additional kilometer per hour of velocity above 80 kph, we expect that the car will use about 0.082 more liters of fuel per hour.

Note that all three of these observations are saying fundamentally the same thing, just from different perspectives: the first regards fuel consumption measured in liters per kilometer; the second regards fuel consumption in kilometers per liter; and the last regards fuel consumption in liters per hour.

2.3 The sine and cosine functions

Motivating Questions

- What is a graphical justification for why $\frac{d}{dx}[a^x] = a^x \ln(a)$?
 - What do the graphs of $y = \sin(x)$ and $y = \cos(x)$ suggest as formulas for their respective derivatives?
 - Once we know the derivatives of $\sin(x)$ and $\cos(x)$, how do previous derivative rules work when these functions are involved?
-

Throughout Chapter 2, we will develop shortcut derivative rules to help us bypass the limit definition and quickly compute $f'(x)$ from a formula for $f(x)$. In Section 2.1, we stated the rule for power functions,

$$\text{if } f(x) = x^n, \text{ then } f'(x) = nx^{n-1},$$

and the rule for exponential functions,

$$\text{if } a \text{ is a positive real number and } f(x) = a^x, \text{ then } f'(x) = a^x \ln(a).$$

Later in this section, we will use a graphical argument to conjecture derivative formulas for the sine and cosine functions.

Example 2.3.1. Consider the function $g(x) = 2^x$, which is graphed in Figure 2.3.2.

- At each of $x = -2, -1, 0, 1, 2$, use a straightedge to sketch an accurate tangent line to $y = g(x)$.
- Use the provided grid to estimate the slope of the tangent line you drew at each point in (a).
- 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).
- 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 $y = g(x)$.
- 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 $g'(x) = 2^x \ln(2)$.

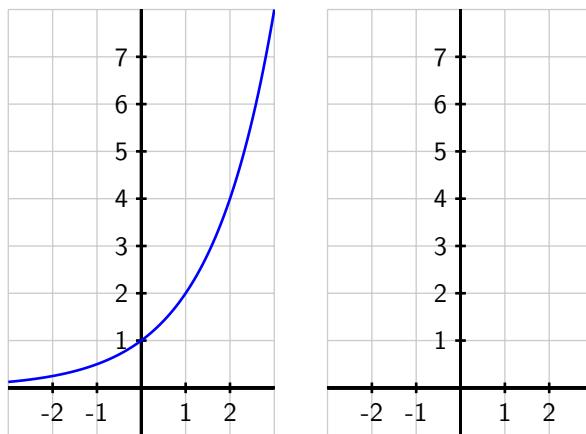


Figure 2.3.2: At left, the graph of $y = g(x) = 2^x$. At right, axes for plotting $y = g'(x)$.

2.3.1 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 definition on the unit circle, these periodic functions play a key role in modeling repeating phenomena such as tidal elevations, the behavior of an oscillating mass attached to a spring, or the location of a point on a bicycle tire. Like polynomial and exponential functions, the sine and cosine functions are considered basic functions, ones that are often used in building 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.

Example 2.3.3. Consider the function $f(x) = \sin(x)$, which is graphed in Figure 2.3.4 below. 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.

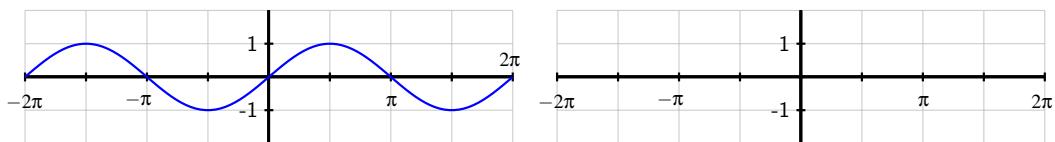


Figure 2.3.4: At left, the graph of $y = f(x) = \sin(x)$.

- 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 $y = f(x)$.
- 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 $f(x) = \sin(x)$?

Hint.

- a. It's very important to use a straightedge for accuracy.
- b. First determine the slopes that appear to be zero. Then estimate $f'(0)$ carefully using the grid. Use symmetry and periodicity to help you estimate other nonzero slopes on the graph.
- c. $f'(0) \approx \frac{\sin(h)}{h}$ for small values of h .
- d. Recall that heights on f' come from slopes on f .
- e. It might be reasonable to expect that the derivative of a trigonometric function is another trigonometric function.

Answer.

- a. Figure 2.3.5.
- b. $1, 0, -1, 0, 1, 0, -1, 0, 1$.
- c. $f'(0) = f'(-2\pi) = f'(2\pi) = 1$.
- d. Figure 2.3.5.
- e. $\frac{d}{dx}[\sin(x)] = \cos(x)$.

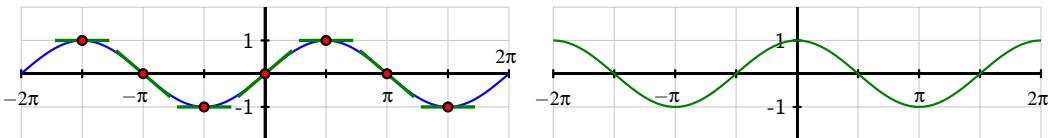


Figure 2.3.5: At left, the graph of $y = f(x) = \sin(x)$. At right, the graph of $y = f'(x)$.

Solution.

- a. See Figure 2.3.6.

- b. Reading left to right from $-2\pi, \dots, 2\pi$ with stepsize $\pi/2$, the respective slopes of tangent lines appear to be $1, 0, -1, 0, 1, 0, -1, 0, 1$.
- c. From the limit definition,

$$\begin{aligned} f'(0) &= \lim_{h \rightarrow 0} \frac{f(0+h) - f(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin(0+h) - \sin(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin(h)}{h} \end{aligned}$$

Because we cannot simplify the fraction $\frac{\sin(h)}{h}$ any further algebraically, we estimate the value of the limit using small values of h . Doing so, it appears that $\lim_{h \rightarrow 0} \frac{\sin(h)}{h} = 1$, and thus $f'(0) = 1$. This matches the estimate generated visually by sketching the tangent line at $(0, f(0))$. Finally, by the periodicity of the sine function, we expect the value of the derivative at 0 to match the derivative value at -2π and 2π .

- d. See the figure below.
- e. It appears that $\frac{d}{dx}[\sin(x)] = \cos(x)$.

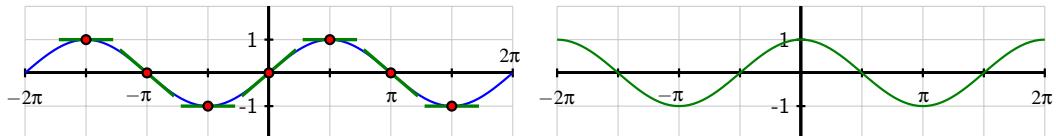


Figure 2.3.6: At left, the graph of $y = f(x) = \sin(x)$. At right, the graph of $y = f'(x)$.

Example 2.3.7. Consider the function $g(x) = \cos(x)$, which is graphed in Figure 2.3.8 below. 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.

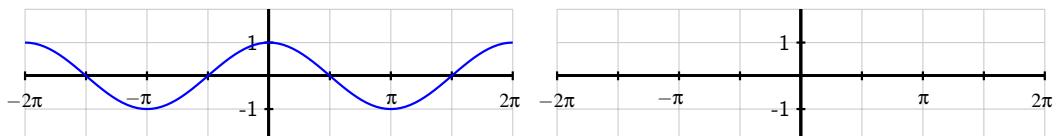


Figure 2.3.8: At left, the graph of $y = g(x) = \cos(x)$.

- 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 $g(x) = \cos(x)$?

Hint.

- a. It's very important to use a straightedge for accuracy.
- b. First determine the slopes that appear to be zero. Then estimate $g'(\frac{\pi}{2})$ carefully using the grid. Use symmetry and periodicity to help you estimate other nonzero slopes on the graph.
- c. $g'(\frac{\pi}{2}) \approx \frac{\cos(\frac{\pi}{2}+h)}{h}$ for small values of h .
- d. Recall that heights on g' come from slopes on g .
- e. It might be reasonable to expect that the derivative of a trigonometric function is another trigonometric function.

Answer.

- a. Figure 2.3.9.
- b. $0, -1, 0, 1, 0, -1, 0, 1, 0$.
- c. $g'(\frac{\pi}{2}) = g'(-\frac{3\pi}{2}) = -1$.
- d. Figure 2.3.9.
- e. $\frac{d}{dx}[\cos(x)] = -\sin(x)$.

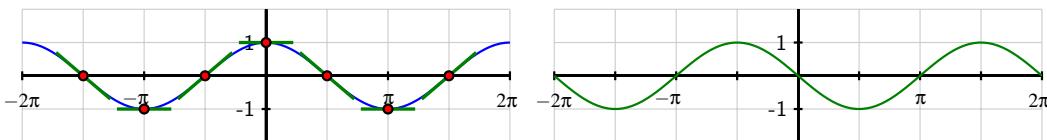


Figure 2.3.9: At left, the graph of $y = g(x) = \cos(x)$. At right, the graph of $y = g'(x)$

Solution.

- a. See Figure 2.3.10..
- b. Reading left to right from $-2\pi, \dots, 2\pi$ with stepsize $\pi/2$, the respective slopes of tangent lines appear to be $0, -1, 0, 1, 0, -1, 0, 1, 0$.
- c. From the limit definition,

$$\begin{aligned} g'\left(\frac{\pi}{2}\right) &= \lim_{h \rightarrow 0} \frac{g\left(\frac{\pi}{2} + h\right) - g\left(\frac{\pi}{2}\right)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\cos\left(\frac{\pi}{2} + h\right) - \cos\left(\frac{\pi}{2}\right)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\cos\left(\frac{\pi}{2}h\right)}{h} \end{aligned}$$

Because we cannot simplify the fraction $\frac{\cos\left(\frac{\pi}{2}h\right)}{h}$ any further algebraically, we estimate the value of the limit using small values of h . Doing so, it appears that $\lim_{h \rightarrow 0} \frac{\cos\left(\frac{\pi}{2}h\right)}{h} = -1$, and thus $g'\left(\frac{\pi}{2}\right) = -1$. This matches the estimate generated visually by sketching the tangent line at $(\frac{\pi}{2}, g(\frac{\pi}{2}))$. Finally, by the periodicity of the sine function, we expect the value of the derivative at $\frac{\pi}{2}$ to match the derivative value at $-\frac{3\pi}{2}$.

- d. See Figure 2.3.9.
- e. It appears that $\frac{d}{dx}[\cos(x)] = -\sin(x)$.

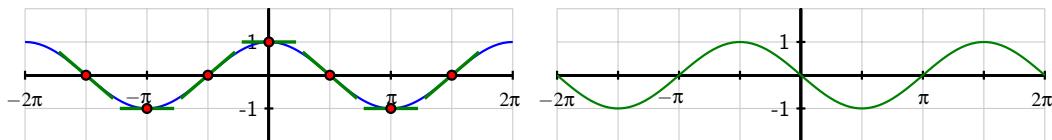


Figure 2.3.10: At left, the graph of $y = g(x) = \cos(x)$. At right, the graph of $y = g'(x)$

The results of the two preceding activities suggest that the sine and cosine functions not only have beautiful 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 activities¹.

Sine and Cosine Functions

For all real numbers x ,

$$\frac{d}{dx}[\sin(x)] = \cos(x) \text{ and } \frac{d}{dx}[\cos(x)] = -\sin(x).$$

We have now added the sine and cosine functions to our library of basic functions whose derivatives we know. The constant multiple and sum rules still hold, of course, as well as all of the inherent meaning of the derivative.

¹These two rules may be formally proved using the limit definition of the derivative and the expansion identities for $\sin(x + h)$ and $\cos(x + h)$.

Example 2.3.11. Answer each of the following questions. Where a derivative is requested, be sure to label the derivative function with its name using proper notation.

- Determine the derivative of $h(t) = 3 \cos(t) - 4 \sin(t)$.
- Find the exact slope of the tangent line to $y = f(x) = 2x + \frac{\sin(x)}{2}$ at the point where $x = \frac{\pi}{6}$.
- Find the equation of the tangent line to $y = g(x) = x^2 + 2 \cos(x)$ at the point where $x = \frac{\pi}{2}$.
- Determine the derivative of $p(z) = z^4 + 4^z + 4 \cos(z) - \sin(\frac{\pi}{2})$.
- 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.

Hint.

- Recall the constant multiple and sum rules.
- $f'(\frac{\pi}{6})$ tells us the slope of the tangent line at $(\frac{\pi}{6}, \frac{\pi}{6})$.
- Find both $(\frac{\pi}{2}, g(\frac{\pi}{2}))$ and $g'(\frac{\pi}{2})$.
- $\sin(\frac{\pi}{2})$ is a constant.
- $P'(a)$ tells us the instantaneous rate of change of P with respect to time at the instant $t = a$, and its units are “units of P per unit of time.”

Answer.

- $\frac{dh}{dt} = -3 \sin(t) - 4 \cos(t)$.
- $y - \frac{\pi^2}{4} = (\pi - 1)(x - \frac{\pi}{2})$.
- $p'(z) = 4z^3 + 4^z \ln(4) - 4 \sin(z)$.
- $P'(2) = 8 \cos(2) \approx -3.329$ hundred animals per decade.

Solution.

- By the sum and constant multiple rules, $\frac{dh}{dt} = 3(-\sin(t)) - 4(\cos(t)) = -3 \sin(t) - 4 \cos(t)$.
- The exact slope of the tangent line to $y = f(x) = 2x + \frac{\sin(x)}{2}$ at $x = \frac{\pi}{6}$ is given by $f'(\frac{\pi}{6})$. So, we first compute $f'(x)$. Using the sum and constant multiple rules, $f'(x) = 2 + \frac{1}{2} \cos(x)$, and thus $f'(\frac{\pi}{6}) = 2 + \frac{1}{2} \cos(\frac{\pi}{6}) = 2 + \frac{\sqrt{3}}{4}$.

- c. The tangent line passes through the point $(\frac{\pi}{2}, g(\frac{\pi}{2}))$ with slope $g'(\frac{\pi}{2})$. We observe first that $g(\frac{\pi}{2}) = (\frac{\pi}{2})^2 + 2 \cos(\frac{\pi}{2}) = \frac{\pi^2}{4}$. Next, we compute the derivative function, $g'(x)$, and find that

$$g'(x) = 2x - 2 \sin(x).$$

Thus, $g(\frac{\pi}{2}) = 2 \cdot \frac{\pi}{2} - 2 \sin(\frac{\pi}{2}) = \pi - 1$. Hence the equation of the tangent line (in point-slope form) is given by

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

- d. Noting that $\sin(\frac{\pi}{2})$ is a constant, we have $p'(z) = 4z^3 + 4^z \ln(4) - 4 \sin(z)$.
- e. The value of $P'(2)$ will tell us the instantaneous rate of change of P at the instant two decades have elapsed. Observe that $P'(t) = 8 \cos(t)$, and thus $P'(2) = 8 \cos(2) \approx -3.329$ hundred animals per decade. This tells us that the instantaneous rate of change of P on January 1, 2030 is about -3329 animals per decade, which tells us that the animal population is shrinking considerably at this point in time. We might say that for whatever the population is on January 1, 2030, we expect that population to drop by about 3300 animals over the next ten years, provided the current population trend continues.

2.3.2 Summary

- For an exponential function $f(x) = a^x$ ($a > 1$), the graph of $f'(x)$ appears to be a scaled version of the original function. In particular, careful analysis of the graph of $f(x) = 2^x$, suggests that $\frac{d}{dx}[2^x] = 2^x \ln(2)$, which is a special case of the rule we stated in Section 2.1.
- By carefully analyzing the graphs of $y = \sin(x)$ and $y = \cos(x)$, and by using the limit definition of the derivative at select points, we found that $\frac{d}{dx}[\sin(x)] = \cos(x)$ and $\frac{d}{dx}[\cos(x)] = -\sin(x)$.
- We note that all previously encountered derivative rules still hold, but now may also be applied to functions involving the sine and cosine. All of the established meaning of the derivative applies to these trigonometric functions as well.

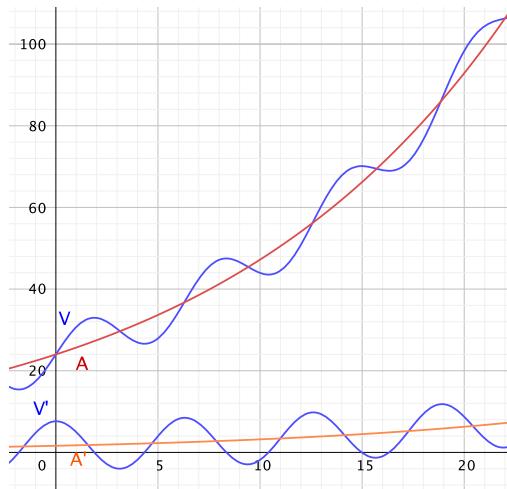
2.3.3 Exercises

- Suppose that $V(t) = 24 \cdot 1.07^t + 6 \sin(t)$ represents the value of a person's investment portfolio in thousands of dollars in year t , where $t = 0$ corresponds to January 1, 2010.
 - At what instantaneous rate is the portfolio's value changing on January 1, 2012? Include units on your answer.
 - Determine the value of $V''(2)$. What are the units on this quantity and what does it tell you about how the portfolio's value is changing?

- c. On the interval $0 \leq t \leq 20$, graph the function $V(t) = 24 \cdot 1.07^t + 6 \sin(t)$ and describe its behavior in the context of the problem. Then, compare the graphs of the functions $A(t) = 24 \cdot 1.07^t$ and $V(t) = 24 \cdot 1.07^t + 6 \sin(t)$, as well as the graphs of their derivatives $A'(t)$ and $V'(t)$. What is the impact of the term $6 \sin(t)$ on the behavior of the function $V(t)$?

Answer.

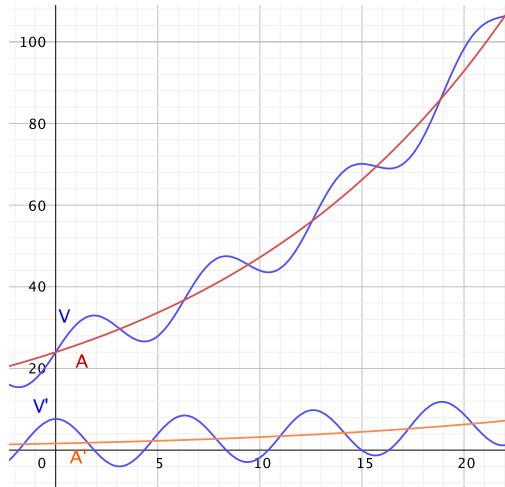
- a. $V'(2) = 24 \cdot 1.07^2 \cdot \ln(1.07) + 6 \cos(2) \approx -0.63778$ thousands of dollars per year.
- b. $V''(2) = 24 \cdot 1.07^2 \cdot \ln(1.07)^2 - 6 \sin(2) \approx -5.33$ thousands of dollars per year per year.
At this moment, V' is decreasing and we expect the derivative's value to decrease by about 5.33 thousand dollars per year over the course of the next year.
- c. See the figure below. Adding the term $6 \sin(t)$ to A to create the function V adds volatility to the value of the portfolio.



Solution.

- a. January 1, 2012 corresponds to the instant $t = 2$, so we compute $V'(2)$. First, we observe by the sum and constant multiple rules that $V'(t) = 24 \cdot 1.07^t \cdot \ln(1.07) + 6 \cos(t)$. Thus, $V'(2) = 24 \cdot 1.07^2 \cdot \ln(1.07) + 6 \cos(2) \approx -0.63778$. Since V is measured in thousands of dollars and t in years, we know that the portfolio's value is decreasing at a rate of 637.78 dollars per year on January 1, 2012.
- b. Since $V'(t) = 24 \cdot 1.07^t \cdot \ln(1.07) + 6 \cos(t)$, it follows that $V''(t) = 24 \cdot 1.07^t \cdot (\ln(1.07))^2 - 6 \sin(t)$, and thus $V''(2) = 24 \cdot 1.07^2 \cdot \ln(1.07)^2 - 6 \sin(2) \approx -5.33$, with units "thousands of dollars per year per year." This means that at the moment $t = 2$, the portfolio's value is decreasing at a decreasing rate. More specifically, we know that V' is decreasing at the moment $t = 2$ and expect the derivative's value to decrease by about 5.33 thousand dollars per year over the course of the next year. In other words, the portfolio's value is not only decreasing, but we expect it to decrease more rapidly as the year goes on.
- c. In the figure below, we see plots of both V and A along with their corresponding derivatives. We see that the function A is a purely exponential function that always increases at an increasing rate. By adding the term $6 \sin(t)$ to A to create the function V , we add volatility to the value of the portfolio: the oscillating nature of the sine function not only

creates variance in the stock's value, but also in its rate of change. We also observe that as time progresses and the value of $24 \cdot (1.07)^t$ increases, the effect of $6 \sin(t)$ lessens.



2. Let $f(x) = 3 \cos(x) - 2 \sin(x) + 6$.

- Determine the exact slope of the tangent line to $y = f(x)$ at the point where $a = \frac{\pi}{4}$.
- Determine the tangent line approximation to $y = f(x)$ at the point where $a = \pi$.
- At the point where $a = \frac{\pi}{2}$, is f increasing, decreasing, or neither?
- At the point where $a = \frac{3\pi}{2}$, does the tangent line to $y = f(x)$ lie above the curve, below the curve, or neither? How can you answer this question without even graphing the function or the tangent line?

Answer.

- $f'(\frac{\pi}{4}) = -5\left(\frac{\sqrt{2}}{2}\right)$.
- $L(x) = 3 + 2(x - \pi)$.
- Decreasing.
- The tangent line to f lies above the curve at this point.

Solution.

- The slope of the line tangent to the graph of f at $a = \frac{\pi}{4}$ is given by $f'(\frac{\pi}{4})$. To find $f'(\frac{\pi}{4})$ we first find $f'(x)$, which is $f'(x) = -3 \sin(x) - 2 \cos(x)$. Thus,

$$\begin{aligned} f'(\frac{\pi}{4}) &= -3 \sin\left(\frac{\pi}{4}\right) - 2 \cos\left(\frac{\pi}{4}\right) \\ &= -3\left(\frac{\sqrt{2}}{2}\right) - 2\left(\frac{\sqrt{2}}{2}\right) = -5\left(\frac{\sqrt{2}}{2}\right). \end{aligned}$$

- The linearization $L(x)$ to f at $a = \pi$ is $L(x) = f(\pi) + f'(\pi)(x - \pi)$. We calculated $f'(x)$ in part (a) and so $f'(\pi) = -3 \sin(\pi) - 2 \cos(\pi) = 2$. We also have $f(\pi) = 3 \cos(\pi) - 2 \sin(\pi) + 6 = 3$. Hence $L(x) = 3 + 2(x - \pi)$.

- c. A differentiable function f is increasing at $x = a$ if and only if $f'(a) > 0$ and decreasing if and only if $f'(a) < 0$. Now $f'(\frac{\pi}{2}) = -3 \sin(\frac{\pi}{2}) - 2 \cos(\frac{\pi}{2}) = -3$, so f is decreasing at $a = \frac{\pi}{2}$.
- d. If a curve defined by a function f is concave up at $x = a$, then the tangent line to f at $x = a$ lies below the curve and if the graph of f is concave down at $x = a$, then the tangent line to f at $x = a$ lies above the curve. The concavity of a curve is determined by the second derivative and $f''(x) = -3 \cos(x) + 2 \sin(x)$. Since

$$f''\left(\frac{3\pi}{2}\right) = -3 \cos\left(\frac{3\pi}{2}\right) + 2 \sin\left(\frac{3\pi}{2}\right) = -2,$$

the graph of f is concave down at $a = \frac{3\pi}{2}$ and the tangent line to f lies above the curve at this point.

3. In this exercise, 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

$$f'(x) = \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin(x)}{h}.$$

- 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

$$f'(x) = \lim_{h \rightarrow 0} \frac{\sin(x)(\cos(h) - 1) + \cos(x)\sin(h)}{h}.$$

- b. Next, note that as h changes, x remains constant. Explain why it therefore makes sense to say that

$$f'(x) = \sin(x) \cdot \lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} + \cos(x) \cdot \lim_{h \rightarrow 0} \frac{\sin(h)}{h}.$$

- c. Finally, use small values of h to estimate the values of the two limits in (c):

$$\lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} \text{ and } \lim_{h \rightarrow 0} \frac{\sin(h)}{h}.$$

- d. What do your results in (b) and (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 $\frac{d}{dx}[\cos(x)] = -\sin(x)$.

Answer.

- a. Hint: in the numerator of the difference quotient, combine the first and last terms and remove a factor of $\sin(x)$.
- b. Hint: divide each part of the numerator by h and consider the sum of two separate limits.
- c. $\lim_{h \rightarrow 0} \left(\frac{\cos(h)-1}{h} \right) = 0$ and $\lim_{h \rightarrow 0} \left(\frac{\sin(h)}{h} \right) = 1$.
- d. $f'(x) = \sin(x) \cdot 0 + \cos(x) \cdot 1$.
- e. Hint: $\cos(\alpha + \beta)$ is $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$.

Solution.

- a. By definition, we know that

$$f'(x) = \lim_{h \rightarrow 0} \frac{\sin(x + h) - \sin(x)}{h}.$$

By the sum of two angles identity for the sine function, $\sin(x + h) = \sin(x)\cos(h) + \cos(x)\sin(h)$, and thus

$$f'(x) = \lim_{h \rightarrow 0} \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h}..$$

Combining the first and last terms and removing a factor of $\sin(x)$,

$$f'(x) = \lim_{h \rightarrow 0} \frac{\sin(x)(\cos(h) - 1) + \cos(x)\sin(h) - \sin(x)}{h}..$$

- b. Continuing our work in (a), we divide each part of the numerator by h and see it follows

$$f'(x) = \lim_{h \rightarrow 0} \left(\frac{\sin(x)(\cos(h) - 1)}{h} + \frac{\cos(x)\sin(h)}{h} \right).$$

Next, we note that $\sin(x)$ does not depend on h , and likewise $\cos(x)$ is independent of h . It thus makes sense to view those constants as being outside the limit being taken, and also that the limit of a sum is the sum of the limits. Hence we write

$$f'(x) = \sin(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\cos(h) - 1}{h} \right) + \cos(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\sin(h)}{h} \right).$$

- c. Using small positive and negative values of h to estimate the desired limits, we find that it appears

$$\lim_{h \rightarrow 0} \left(\frac{\cos(h) - 1}{h} \right) = 0$$

and

$$\lim_{h \rightarrow 0} \left(\frac{\sin(h)}{h} \right) = 1$$

It suffices in each case to consider h -values of $h = \pm 0.01, \pm 0.001, \pm 0.0001$.

- d. Combining our results in (b) and (c), we have found that

$$f'(x) = \sin(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\cos(h) - 1}{h} \right) + \cos(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\sin(h)}{h} \right) = \sin(x) \cdot 0 + \cos(x) \cdot 1 = \cos(x).$$

- e. The argument for showing $\frac{d}{dx}[\cos(x)] = -\sin(x)$ is almost identical. Do note that the sum of two angles identity for $\cos(\alpha + \beta)$ is $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$.

2.4 Derivatives of other trigonometric functions

Motivating Questions

- 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?

One of the powerful themes in trigonometry comes from a very simple idea: locating a point on the unit circle.

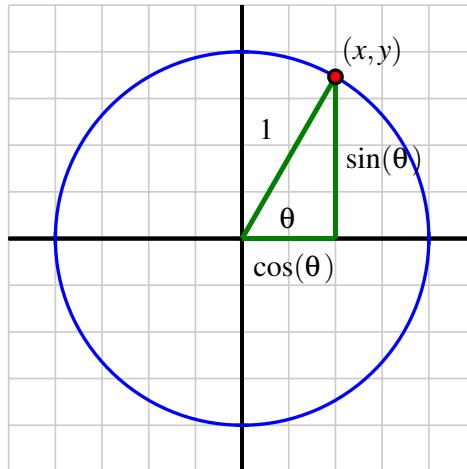


Figure 2.4.1: The unit circle and the definition of the sine and cosine functions.

Because each angle θ in standard position corresponds to one and only one point (x, y) on the unit circle, the x - and y -coordinates of this point are each functions of θ . In fact, this is the very definition of $\cos(\theta)$ and $\sin(\theta)$: $\cos(\theta)$ is the x -coordinate of the point on the unit circle corresponding to the angle θ , and $\sin(\theta)$ is the y -coordinate. From this simple definition, all of trigonometry is founded. For instance, the Fundamental Trigonometric Identity,

$$\sin^2(\theta) + \cos^2(\theta) = 1,$$

is a restatement of the Pythagorean Theorem, applied to the right triangle shown in Figure 2.4.1.

There are four other trigonometric functions, each defined in terms of the sine and/or cosine functions.

- The tangent function is defined by $\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$;
- the cotangent function is its reciprocal: $\cot(\theta) = \frac{\cos(\theta)}{\sin(\theta)}$.
- The secant function is the reciprocal of the cosine function, $\sec(\theta) = \frac{1}{\cos(\theta)}$;
- and the cosecant function is the reciprocal of the sine function, $\csc(\theta) = \frac{1}{\sin(\theta)}$.

These six trigonometric functions together offer us a wide range of flexibility in problems involving right triangles.

Because we know the derivatives of the sine and cosine function, we can now develop short-cut differentiation rules for the tangent, cotangent, secant, and cosecant functions. In Example 2.4.2, we work through the steps to find the derivative of $y = \tan(x)$.

Example 2.4.2. Consider the function $f(x) = \tan(x)$, and remember that $\tan(x) = \frac{\sin(x)}{\cos(x)}$.

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 $\sec(x) = \frac{1}{\cos(x)}$. 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 ?

2.4.1 Derivatives of the cotangent, secant, and cosecant functions

In Example 2.4.2, 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 $g'(x) = -\frac{1}{\sin^2(x)}$, and recalling that $\csc(x) = \frac{1}{\sin(x)}$, it follows that we can 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 $k = 0, \pm 1, \pm 2, \dots$,

$$\frac{d}{dx}[\cot(x)] = -\csc^2(x).$$

Notice that the derivative of the cotangent function is very similar to the derivative of the tangent function we discovered in Example 2.4.2.

Tangent Function

For all real numbers x such that $x \neq \frac{(2k+1)\pi}{2}$, where $k = \pm 1, \pm 2, \dots$,

$$\frac{d}{dx}[\tan(x)] = \sec^2(x).$$

In the next two activities, we develop the rules for differentiating the secant and cosecant functions.

Example 2.4.3. Let $h(x) = \sec(x)$ and recall that $\sec(x) = \frac{1}{\cos(x)}$.

- What is the domain of h ?
- Use the quotient rule to develop a formula for $h'(x)$ that is expressed completely in terms of $\sin(x)$ and $\cos(x)$.
- How can you use other relationships among trigonometric functions to write $h'(x)$ only in terms of $\tan(x)$ and $\sec(x)$?
- What is the domain of h' ? How does this compare to the domain of h ?

Hint.

- For what values of x is $\cos(x) = 0$?
- Don't forget that $\frac{d}{dx}[1] = 0$.
- Consider rewriting $\frac{\sin(x)}{\cos^2(x)}$ as $\frac{1}{\cos(x)} \cdot \frac{\sin(x)}{\cos(x)}$.
- Observe that $\cos(x)$ is still present in the denominator of $h'(x)$.

Answer.

- All real numbers x such that $x \neq \frac{k\pi}{2}$, where $k = \pm 1, \pm 2, \dots$

- $$h'(x) = \frac{\sin(x)}{\cos^2(x)}.$$

- c. $h'(x) = \sec(x) \tan(x)$.
- d. h and h' have the same domain: all real numbers x such that $x \neq \frac{k\pi}{2}$, where $k = \pm 1, \pm 2, \dots$

Solution.

- a. $h(x) = \sec(x)$ is defined for all x for which $\cos(x) \neq 0$. Hence the domain of h is all real numbers x such that $x \neq \frac{k\pi}{2}$, where $k = \pm 1, \pm 2, \dots$

- b. By the quotient rule,

$$h'(x) = \frac{0 - 1(-\sin(x))}{\cos^2(x)} = \frac{\sin(x)}{\cos^2(x)}.$$

- c. Observe that $h'(x) = \frac{\sin(x)}{\cos^2(x)} = \frac{1}{\cos(x)} \cdot \frac{\sin(x)}{\cos(x)}$, so

$$h'(x) = \sec(x) \tan(x).$$

- d. The derivative $h'(x)$ is, like $h(x)$, defined for all values of x for which $\cos(x) \neq 0$. Therefore, h and h' have the same domain: all real numbers x such that $x \neq \frac{k\pi}{2}$, where $k = \pm 1, \pm 2, \dots$

Example 2.4.4. Let $p(x) = \csc(x)$ and recall that $\csc(x) = \frac{1}{\sin(x)}$.

- 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 ?

Hint.

- a. For what values of x is $\sin(x) = 0$?
- b. Don't forget that $\frac{d}{dx}[1] = 0$.
- c. Consider rewriting $\frac{\cos(x)}{\sin^2(x)}$ as $\frac{1}{\sin(x)} \cdot \frac{\cos(x)}{\sin(x)}$.
- d. Observe that $\sin(x)$ is still present in the denominator of $h'(x)$.

Answer.

- a. All real numbers x such that $xk\pi$, where $k = 0, \pm 1, \pm 2, \dots$
- b. $h'(x) = \frac{\cos(x)}{\sin^2(x)}$.

- c. $h'(x) = -\csc(x)\cot(x)$.
- d. p and p' have the same domain: all real numbers x such that $x \neq k\pi$, where $k = 0, \pm 1, \pm 2, \dots$

Solution.

- a. $p(x) = \csc(x)$ is defined for all x for which $\sin(x) \neq 0$. Hence the domain of h is all real numbers x such that $x \neq k\pi$, where $k = 0, \pm 1, \pm 2, \dots$

- b. By the quotient rule,

$$h'(x) = \frac{0 - 1 \cdot (\cos(x))}{\sin^2(x)} = -\frac{\cos(x)}{\sin^2(x)}.$$

- c. Observe that $h'(x) = -\frac{\cos(x)}{\sin^2(x)} = -\frac{1}{\sin(x)} \cdot \frac{\cos(x)}{\sin(x)}$, so

$$h'(x) = -\csc(x)\cot(x).$$

- d. The derivative $p'(x)$ is, like $p(x)$, defined for all values of x for which $\sin(x) \neq 0$. Therefore, p and p' have the same domain: all real numbers x such that $x \neq k\pi$, where $k = 0, \pm 1, \pm 2, \dots$

Using the quotient rule we have determined the derivatives of the tangent, cotangent, secant, and cosecant functions, expanding our overall library of functions we can differentiate. 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.

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.

Example 2.4.5. 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 $h(t) = \frac{\tan(t)}{t^2 + 1} - 2e^t \cos(t)$. Find $h'(t)$.
- d. Let $g(r) = \frac{r \sec(r)}{5^r}$. 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 of this function, and explain what it tells you about the object's motion. Finally, compute and interpret $s'(2)$.

Hint.

- What rule(s) can help you determine $f'(x)$?
- Note that p is a sum of two functions. What rule is needed to differentiate each term in the sum?
- Observe that h is a sum of two functions; the first term in the sum is a quotient, while the second is a product.
- What is the overall structure of g ? What is the algebraic structure of the numerator of g ?
- Keep in mind that the derivative of position tells us the instantaneous velocity.

Answer.

a. $m = f'(\frac{\pi}{3}) = 10\sqrt{3} + \frac{4}{3}$.

b. $p'(\frac{\pi}{4}) = \frac{\pi^2}{16} \sqrt{2} + \frac{\sqrt{2}\pi}{2} + \frac{\pi}{2} - 1$.

c. $h'(t) = \frac{(t^2+1)\sec^2(t)-2t\tan(t)}{(t^2+1)^2} + 2e^t \sin(t) - 2e^t \cos(t)$.

d. $g'(r) = \frac{r \sec(r) \tan(r) + \sec(r) - r^5 \sec(r)}{5r}$.

e. $s'(2) = \frac{15 \cos(2) - 15 \sin(2)}{e^2} \approx -2.69$ inches per second.

Solution.

- Using the sum and constant multiple rules along with the formulas for the derivatives of $\sec(x)$ and $\csc(x)$, we find that

$$f'(x) = 5 \sec(x) \tan(x) + 2 \csc(x) \cot(x).$$

Therefore, the slope of the tangent line to f at the point where $x = \frac{\pi}{3}$ is given by

$$\begin{aligned} m &= f'(\frac{\pi}{3}) = 5 \sec(\frac{\pi}{3}) \tan(\frac{\pi}{3}) + 2 \csc(\frac{\pi}{3}) \cot(\frac{\pi}{3}) \\ &= 5 \cdot 2 \cdot \sqrt{3} + 2 \cdot \frac{2}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} = 10\sqrt{3} + \frac{4}{3}. \end{aligned}$$

- b. By the sum rule and two applications of the product rule, we have

$$\begin{aligned} p'(z) &= \frac{d}{dz}[z^2 \sec(z)] - \frac{d}{dz}[z \cot(z)] \\ &= [z^2 \sec(z) \tan(z) + \sec(z) \cdot 2z] - [z(-\csc^2(z)) + \cot(z) \cdot 1] \\ &= z^2 \sec(z) \tan(z) + 2z \sec(z) + z \csc^2(z) - \cot(z). \end{aligned}$$

Thus, the instantaneous rate of change of p at the point where $z = \frac{\pi}{4}$ is

$$\begin{aligned} p'\left(\frac{\pi}{4}\right) &= \left(\frac{\pi}{4}\right)^2 \sec\left(\frac{\pi}{4}\right) \tan\left(\frac{\pi}{4}\right) + 2\frac{\pi}{4} \sec\left(\frac{\pi}{4}\right) + \frac{\pi}{4} \csc^2\left(\frac{\pi}{4}\right) - \cot\left(\frac{\pi}{4}\right) \\ &= \frac{\pi^2}{16} \sqrt{2} + 2\frac{\pi}{4} \sqrt{2} + \frac{\pi}{4} 2 - 1 \\ &= \frac{\pi^2}{16} \sqrt{2} + \frac{\sqrt{2}\pi}{2} + \frac{\pi}{2} - 1 \end{aligned}$$

- c. Using the sum and constant multiple rules, followed by the quotient rule on the first term and the product rule on the second, we find that

$$\begin{aligned} h'(t) &= \frac{d}{dt} \left[\frac{\tan(t)}{t^2 + 1} \right] - 2 \frac{d}{dt} [e^t \cos(t)] \\ &= \frac{(t^2 + 1) \sec^2(t) - \tan(t)(2t)}{(t^2 + 1)^2} - 2(e^t(-\sin(t)) + \cos(t)e^t) \\ &= \frac{(t^2 + 1) \sec^2(t) - 2t \tan(t)}{(t^2 + 1)^2} + 2e^t \sin(t) - 2e^t \cos(t) \end{aligned}$$

- d. Note that g is fundamentally a quotient, so we need to use the quotient rule. But the numerator of g is a product, so the product rule will be required to compute the derivative of the top function. Executing the quotient rule and proceeding, we find that

$$\begin{aligned} g'(r) &= \frac{5^r \frac{d}{dr}[r \sec(r)] - r \sec(r) \cdot 5^r \ln(5)}{(5^r)^2} \\ &= \frac{5^r[r \sec(r) \tan(r) + \sec(r) \cdot 1] - r \sec(r) \cdot 5^r \ln(5)}{(5^r)^2} \\ &= \frac{r \sec(r) \tan(r) + \sec(r) - r 5^r \sec(r)}{5^r} \end{aligned}$$

- e. By the quotient rule,

$$\frac{ds}{dt} = \frac{e^t \cdot 15 \cos(t) - 15 \sin(t) \cdot e^t}{(e^t)^2} = \frac{15 \cos(t) - 15 \sin(t)}{e^t}.$$

The function $\frac{ds}{dt} = s'(t)$ measures the instantaneous vertical velocity of the mass that is attached to the spring. In particular, $s'(2) = \frac{15 \cos(2) - 15 \sin(2)}{e^2} \approx -2.69$ inches per second, which tells us at the instant $t = 2$, the mass is moving downward at an instantaneous rate of 2.69 inches per second.

2.4.2 Summary

- 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 $k = \pm 1, \pm 2, \dots$

- The 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.

2.4.3 Exercises

1. (A sum and product involving $\tan(x)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (A quotient involving $\tan(t)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (A quotient of trigonometric functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (A quotient that involves a product) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Finding a tangent line equation) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. An object moving vertically has its height at time t (measured in feet, with time in seconds) given by the function $h(t) = 3 + \frac{2\cos(t)}{1.2^t}$.
 - What is the object's instantaneous velocity when $t = 2$?
 - What is the object's acceleration at the instant $t = 2$?
 - Describe in everyday language the behavior of the object at the instant $t = 2$.

Answer.

a. $h'(2) = \frac{-2\sin(2)-2\cos(2)\ln(1.2)}{1.2^2} \approx -1.1575$ feet per second.

b. $h''(2) = \frac{\cos(2)(-2+2\ln^2(1.2))+4\ln(1.2)\sin(2)}{1.2^2} \approx 1.0193$ feet per second per second.

c. The object is falling and slowing down.

Solution.

- a. The object's instantaneous velocity at $t = 2$ is given by $h'(2)$. Since $h(t) = 3 + \frac{2\cos(t)}{1.2^t}$, by the quotient rule we can determine that

$$\begin{aligned} h'(t) &= \frac{1.2^t \cdot (-2\sin(t)) - 2\cos(t) \cdot 1.2^t \ln(1.2)}{(1.2^t)^2} \\ &= \frac{1.2^t(-2\sin(t) - 2\cos(t)\ln(1.2))}{(1.2^t)^2} \\ &= \frac{-2\sin(t) - 2\cos(t)\ln(1.2)}{1.2^t}. \end{aligned}$$

Thus, $h'(2) = \frac{-2\sin(2) - 2\cos(2)\ln(1.2)}{1.2^2} \approx -1.1575$ feet per second is the instantaneous (vertical) velocity of the object at $t = 2$.

- b. To find the object's acceleration at $t = 2$, we compute $h''(2)$. First, by the quotient rule,

$$\begin{aligned} h''(t) &= \frac{1.2^t(-2\cos(t) + 2\sin(t)\ln(1.2)) - (-2\sin(t) - 2\cos(t)\ln(1.2))1.2^t \ln(1.2)}{(1.2^t)^2} \\ &= \frac{-2\cos(t) + 2\sin(t)\ln(1.2) + 2\sin(t)\ln(1.2) + 2\cos(t)\ln^2(1.2)}{1.2^t} \\ &= \frac{\cos(t)(-2 + 2\ln^2(1.2)) + 4\ln(1.2)\sin(t)}{1.2^t}. \end{aligned}$$

Therefore, the acceleration at $t = 2$ is $h''(2) = \frac{\cos(2)(-2 + 2\ln^2(1.2)) + 4\ln(1.2)\sin(2)}{1.2^2} \approx 1.0193$ feet per second per second.

- c. At $t = 2$, the object is falling and slowing down. We know the object is falling because its velocity is negative, and slowing down because its negative velocity is increasing (due to $h''(2)$ being positive) and thus getting less negative.

7. Let $f(x) = \sin(x)\cot(x)$.

- Use the product rule to find $f'(x)$.
- True or false: for all real numbers x , $f(x) = \cos(x)$.
- 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' .)

Answer.

- $f'(x) = \sin(x) \cdot (-\csc^2(x)) + \cot(x) \cdot \cos(x)$.
- False.
- $f'(x) = \frac{-\sin^2(x)}{\sin(x)} = -\sin(x)$ for $x \neq \frac{\pi}{2} + k\pi$ for some integer value of k .

Solution.

- Since $f(x) = \sin(x)\cot(x)$, by the product rule we know that $f'(x) = \sin(x) \cdot (-\csc^2(x)) + \cot(x) \cdot \cos(x)$.
- False. Since $\cot(x) = \frac{\cos(x)}{\sin(x)}$, we can write $f(x) = \sin(x) \cdot \frac{\cos(x)}{\sin(x)}$. From here, it is tempting to "cancel" the $\sin(x)$ term in the numerator and denominator. However, note that

because of the $\sin(x)$ in the denominator, f is not defined for $x = \frac{\pi}{2} + k\pi$ for any integer k . In other words, the domain of f is all real numbers except odd multiples of $\frac{\pi}{2}$. Thus, we can say that $f(x) = \cos(x)$ for all real numbers except odd multiples of $\frac{\pi}{2}$. But it is not the case that $f(x) = \cos(x)$ for every value of x .

- c. Taking our result from (a) and converting $\csc(x)$ and $\cot(x)$ to expressions involving the sine and cosine functions, we have

$$\begin{aligned} f'(x) &= \sin(x) \cdot (-\csc^2(x)) + \cot(x) \cdot \cos(x) \\ &= \sin(x) \cdot \left(-\frac{1}{\sin^2(x)}\right) + \frac{\cos(x)}{\sin(x)} \cdot \cos(x) \end{aligned}$$

Provided that $x \neq \frac{\pi}{2} + k\pi$ for some integer value of k , it follows that

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

By the Fundamental Trigonometric Identity, recall that $\cos^2(x) + \sin^2(x) = 1$, and thus $\cos^2(x) - 1 = -\sin^2(x)$. Hence, for $x \neq \frac{\pi}{2} + k\pi$ for some integer value of k , we have $f'(x) = \frac{-\sin^2(x)}{\sin(x)} = -\sin(x)$. Again, this equality holds only for $x \neq \frac{\pi}{2} + k\pi$ for some integer value of k because those values of x are not in the domain of the original function f .

8. Let $p(z)$ be given by the rule

$$p(z) = \frac{z \tan(z)}{z^2 \sec(z) + 1} + 3e^z + 1.$$

- Determine $p'(z)$.
- Find an equation for the tangent line to p at the point where $z = 0$.
- At $z = 0$, is p increasing, decreasing, or neither? Why?

Answer.

- $p'(z) = \frac{(z^2 \sec(z) + 1)(z \sec^2(z) + \tan(z)) - z \tan(z)(z^2 \sec(z) \tan(z) + 2z \sec(z))}{(z^2 \sec(z) + 1)^2} + 3e^z$
- $y - 4 = 3(x - 0)$.
- Increasing.

Solution.

- We use the quotient, product, sum, and constant multiple rules to obtain

$$\begin{aligned} p'(z) &= \frac{d}{dz} \left(\frac{z \tan(z)}{z^2 \sec(z) + 1} + 3e^z + 1 \right) \\ &= \frac{d}{dz} \left(\frac{z \tan(z)}{z^2 \sec(z) + 1} \right) + \frac{d}{dz} (3e^z) + \frac{d}{dz} (1) \\ &= \frac{(z^2 \sec(z) + 1) \frac{d}{dz} (z \tan(z)) - z \tan(z) \frac{d}{dz} (z^2 \sec(z) + 1)}{(z^2 \sec(z) + 1)^2} + 3e^z \\ &= \frac{(z^2 \sec(z) + 1)(z \sec^2(z) + \tan(z)) - z \tan(z)(z^2 \sec(z) \tan(z) + 2z \sec(z))}{(z^2 \sec(z) + 1)^2} + 3e^z \end{aligned}$$

- b. We have $p(0) = 0 + 3 + 1 = 4$ and $p'(0) = 0 + 3 = 3$, so an equation for the tangent line to p at the point where $z = 0$ is $y - 4 = 3(x - 0)$.
- c. We found that $p'(0) = 3$ is positive, so p is increasing at $t = 0$.

2.5 The chain rule

Motivating 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 $C'(x)$ in terms of f , g , f' , and g' ? What is the statement of the Chain Rule?
-

In addition to learning how to differentiate a variety of basic functions, we have also been developing our ability to use rules to differentiate certain algebraic combinations of them.

Example 2.5.1. State the rule(s) to find the derivative of each of the following combinations of $f(x) = \sin(x)$ and $g(x) = x^2$:

$$s(x) = 3x^2 - 5\sin(x),$$

$$p(x) = x^2 \sin(x), \text{ and}$$

$$q(x) = \frac{\sin(x)}{x^2}.$$

Solution. Finding s' uses the sum and constant multiple rules, because $s(x) = 3g(x) - 5f(x)$. Determining p' requires the product rule, because $p(x) = g(x) \cdot f(x)$. To calculate q' we use the quotient rule, because $q(x) = \frac{f(x)}{g(x)}$.

There is one more natural way to combine basic functions algebraically, 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 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 the input for the function g , and the result is used as the input for f . We write

$$C(x) = f(g(x)) = \sin(x^2)$$

and say that C is the *composition* of f and g . We will refer to g , the function that is first applied to x , as the *inner* function, while f , the function that is applied to the result, is the *outer* function.

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 is 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 inner and outer functions, as we did in the example above with $C(x) = \sin(x^2)$.

Example 2.5.2. 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 inner function g and the outer function f 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. $h(x) = \tan(2^x)$

d. $m(x) = e^{\tan(x)}$

b. $p(x) = 2^x \tan(x)$

e. $w(x) = \sqrt{x} + \tan(x)$

c. $r(x) = (\tan(x))^2$

f. $z(x) = \sqrt{\tan(x)}$

2.5.1 The chain rule

Often a composite function cannot be written in an alternate algebraic form. For instance, the function $C(x) = \sin(x^2)$ cannot be expanded or otherwise rewritten, so it presents no alternate approaches to taking the derivative. But some composite functions can be expanded or simplified, and these provide a way to explore how the chain rule works.

Example 2.5.3. Let $f(x) = -4x + 7$ and $g(x) = 3x - 5$. Determine a formula for $C(x) = f(g(x))$ and compute $C'(x)$. How is C' related to f and g and their derivatives?

Solution. By the rules given for f and g ,

$$\begin{aligned} C(x) &= f(g(x)) \\ &= f(3x - 5) \\ &= -4(3x - 5) + 7 \\ &= -12x + 20 + 7 \\ &= -12x + 27. \end{aligned}$$

Thus, $C'(x) = -12$. Noting that $f'(x) = -4$ and $g'(x) = 3$, we observe that C' appears to be the product of f' and g' .

It may seem that Example 2.5.3 is too elementary to illustrate how to differentiate a composite function. Linear functions are the simplest of all functions, and composing linear functions yields another linear function. While this example does not illustrate the full complexity of a composition of nonlinear functions, at the same time we remember that any differentiable function is *locally* linear, and thus any function with a derivative behaves like a line when

viewed up close. The fact that the derivatives of the linear functions f and g are multiplied to find the derivative of their composition turns out to be a key insight.

We now consider a composition involving a nonlinear function.

Example 2.5.4. 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. Using the double angle identity for the sine function, we write

$$C(x) = \sin(2x) = 2 \sin(x) \cos(x).$$

Applying the product rule and simplifying, we find

$$C'(x) = 2 \sin(x)(-\sin(x)) + \cos(x)(2 \cos(x)) = 2(\cos^2(x) - \sin^2(x)).$$

Next, we recall that the double angle identity for the cosine,

$$\cos(2x) = \cos^2(x) - \sin^2(x).$$

Substituting this result into our expression for $C'(x)$, we now have that

$$C'(x) = 2 \cos(2x).$$

In Example 2.5.4, if we let $g(x) = 2x$ and $f(x) = \sin(x)$, we observe that $C(x) = f(g(x))$. Now, $g'(x) = 2$ and $f'(x) = \cos(x)$, so we can view the structure of $C'(x)$ as

$$C'(x) = 2 \cos(2x) = g'(x)f'(g(x)).$$

In this example, as in the example involving linear functions, we see that the derivative of the composite function $C(x) = f(g(x))$ is found by multiplying the derivatives of f and g , but with f' evaluated at $g(x)$.

It makes sense intuitively that these two quantities are involved in the rate of change of a composite function: if we ask how fast C is changing at a given x value, it clearly matters how fast g is changing at x , as well as how fast f is changing at the value of $g(x)$. It turns out that this structure holds for all differentiable functions¹ as is stated in the Chain Rule.

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

$$C'(x) = f'(g(x))g'(x).$$

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

¹Like other differentiation rules, the Chain Rule can be proved formally using the limit definition of the derivative.

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."

It is helpful to identify clearly the inner function g and outer function f , compute their derivatives individually, and then put all of the pieces together by the chain rule.

Example 2.5.5. Determine the derivative of the function

$$r(x) = (\tan(x))^2.$$

Solution. 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, we find that

$$r'(x) = f'(g(x))g'(x) = 2 \tan(x) \sec^2(x).$$

As a side note, we remark that $r(x)$ is usually written as $\tan^2(x)$. 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 2.5.6. For each function given below, identify an inner function g and outer function f to write the function in the form $f(g(x))$. Determine $f'(x)$, $g'(x)$, and $f'(g(x))$, and then apply the chain rule to determine the derivative of the given function.

a. $h(x) = \cos(x^4)$

d. $z(x) = \cot^5(x)$

b. $p(x) = \sqrt{\tan(x)}$

c. $s(x) = 2^{\sin(x)}$

e. $m(x) = (\sec(x) + e^x)^9$

Hint.

a. The outer function is $f(x) = \cos(x)$.

b. The outer function is $f(x) = \sqrt{x}$.

c. The outer function is $f(x) = 2^x$.

d. The outer function is $f(x) = x^5$.

e. The outer function is $f(x) = x^9$.

Answer.

a. $h'(x) = -4x^3 \sin(x^4)$.

- b. $h'(x) = \frac{\sec^2(x)}{2\sqrt{\tan(x)}}.$
- c. $h'(x) = 2^{\sin(x)} \ln(2) \cos(x).$
- d. $h'(x) = -5 \cot^4(x) \csc^2(x).$
- e. $h'(x) = 9(\sec(x) + e^x)^8(\sec(x) \tan(x) + e^x).$

Solution.

- a. The outer function is $f(x) = \cos(x)$, while the inner function is $g(x) = x^4$, and we know that

$$f'(x) = -\sin(x), g'(x) = 4x^3, \text{ and } f'(g(x)) = -\sin(x^4).$$

Hence, by the chain rule,

$$h'(x) = f'(g(x))g'(x) = -4x^3 \sin(x^4).$$

- b. The outer function is $f(x) = \sqrt{x}$, while the inner function is $g(x) = \tan(x)$, and we know that

$$f'(x) = \frac{1}{2\sqrt{x}}, g'(x) = \sec^2(x), \text{ and } f'(g(x)) = \frac{1}{2\sqrt{\tan(x)}}.$$

Hence, by the chain rule,

$$h'(x) = f'(g(x))g'(x) = \frac{\sec^2(x)}{2\sqrt{\tan(x)}}.$$

- c. The outer function is $f(x) = 2^x$, while the inner function is $g(x) = \sin(x)$, and we know that

$$f'(x) = 2^x \ln(2), g'(x) = \cos(x), \text{ and } f'(g(x)) = 2^{\sin(x)} \ln(2).$$

Hence, by the chain rule,

$$h'(x) = f'(g(x))g'(x) = 2^{\sin(x)} \ln(2) \cos(x).$$

- d. The outer function is $f(x) = x^5$, while the inner function is $g(x) = \cot(x)$, and we know that

$$f'(x) = 5x^4, g'(x) = -\csc^2(x), \text{ and } f'(g(x)) = 5 \cot^4(x).$$

Hence, by the chain rule,

$$h'(x) = f'(g(x))g'(x) = -5 \cot^4(x) \csc^2(x).$$

- e. The outer function is $f(x) = x^9$, while the inner function is $g(x) = \sec(x) + e^x$, and we know that

$$f'(x) = 9x^8, g'(x) = \sec(x) \tan(x) + e^x, \text{ and } f'(g(x)) = 9(\sec(x) + e^x)^8.$$

Hence, by the chain rule,

$$h'(x) = f'(g(x))g'(x) = 9(\sec(x) + e^x)^8(\sec(x) \tan(x) + e^x).$$

2.5.2 Using multiple rules simultaneously

The chain rule now joins the sum, constant multiple, product, and quotient rules in our collection of techniques for finding the derivative of a function through understanding its algebraic structure and the basic functions that constitute it. It takes practice to get comfortable applying multiple rules to differentiate a single function, but using proper notation and taking a few extra steps will help.

Example 2.5.7. Find a formula for the derivative of $h(t) = 3^{t^2+2t} \sec^4(t)$.

Solution. We first observe that h is the product of two functions: $h(t) = a(t) \cdot b(t)$, where $a(t) = 3^{t^2+2t}$ and $b(t) = \sec^4(t)$. We will need to use the product rule to differentiate h . And because a and b are composite functions, we will need the chain rule. We therefore begin by computing $a'(t)$ and $b'(t)$.

Writing $a(t) = f(g(t)) = 3^{t^2+2t}$, and finding the derivatives of f and g , we have

$$\begin{aligned} f(t) &= 3^t & g(t) &= t^2 + 2t \\ f'(t) &= 3^t \ln(3) & g'(t) &= 2t + 2 \\ f'(g(t)) &= 3^{t^2+2t} \ln(3) \end{aligned}$$

Thus, by the chain rule, it follows that $a'(t) = f'(g(t))g'(t) = 3^{t^2+2t} \ln(3)(2t + 2)$.

Turning next to b , we write $b(t) = r(s(t)) = \sec^4(t)$ and find the derivatives of r and s .

$$\begin{aligned} r(t) &= t^4 & s(t) &= \sec(t) \\ r'(t) &= 4t^3 & s'(t) &= \sec(t) \tan(t) \\ r'(s(t)) &= 4 \sec^3(t) \end{aligned}$$

By the chain rule,

$$b'(t) = r'(s(t))s'(t) = 4 \sec^3(t) \sec(t) \tan(t) = 4 \sec^4(t) \tan(t).$$

Now we are finally ready to compute the derivative of the 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 2.5.8. 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. $h(y) = \frac{\cos(10y)}{e^{4y} + 1}$
- d. $s(z) = 2^{z^2 \sec(z)}$
- e. $c(x) = \sin(e^{x^2})$

Hint.

- Use the constant multiple rule first, followed by the chain rule.
- Observe that m is fundamentally a product of composite functions.
- Note that h is a quotient of composite functions.
- The function s is a composite function with outer function 2^z .
- It is possible for a function to be a composite function with more than two functions in the chain.

Answer.

- a. $p'(r) = \frac{4(6r^5 + 2e^r)}{2\sqrt{r^6 + 2e^r}}.$
- b. $m'(v) = -3v^2 \sin(v^2) \sin(v^3) + 2v \cos(v^3) \cos(v^2).$
- c. $h'(y) = \frac{(e^{4y} + 1)[-10 \sin(10y)] - \cos(10y)[4e^{4y}]}{(e^{4y} + 1)^2}.$
- d. $s'(z) = 2^{z^2 \sec(z)} \ln(2)[z^2 \sec(z) \tan(z) + \sec(z) \cdot 2z].$
- e. $c'(x) = \cos(e^{x^2})[e^{x^2} \cdot 2x].$

Solution.

- a. By the constant multiple rule, $p'(r) = 4 \frac{d}{dr} [\sqrt{r^6 + 2e^r}]$. Using the chain rule to complete the remaining derivative, we see that

$$p'(r) = 4 \frac{1}{2\sqrt{r^6 + 2e^r}} \frac{d}{dr} [r^6 + 2e^r] = \frac{4(6r^5 + 2e^r)}{2\sqrt{r^6 + 2e^r}}.$$

- b. Observe that by the product rule, $m'(v) = \sin(v^2) \frac{d}{dv} [\cos(v^3)] + \cos(v^3) \frac{d}{dv} [\sin(v^2)]$. Applying the chain rule to differentiate $\cos(v^3)$ and $\sin(v^2)$, we see that

$$m'(v) = \sin(v^2)[- \sin(v^3) \cdot 3v^2] + \cos(v^3)[\cos(v^2) \cdot 2v] = -3v^2 \sin(v^2) \sin(v^3) + 2v \cos(v^3) \cos(v^2).$$

- c. By the quotient rule,

$$h'(y) = \frac{(e^{4y} + 1) \frac{d}{dy} [\cos(10y)] - \cos(10y) \frac{d}{dy} [e^{4y} + 1]}{(e^{4y} + 1)^2}.$$

Applying the chain rule to differentiate $\cos(10y)$ and e^{4y} , it follows

$$h'(y) = \frac{(e^{4y} + 1)[-10 \sin(10y)] - \cos(10y)[4e^{4y}]}{(e^{4y} + 1)^2}.$$

- d. By the chain rule, $s'(z) = 2^{z^2 \sec(z)} \ln(2) \frac{d}{dz} [z^2 \sec(z)]$. Then by the product rule, we find that

$$s'(z) = 2^{z^2 \sec(z)} \ln(2) [z^2 \sec(z) \tan(z) + \sec(z) \cdot 2z].$$

- e. If we first apply the chain rule to the outer function (the sine function), note that

$$c'(x) = \cos(e^{x^2}) \frac{d}{dx} [e^{x^2}].$$

Next, we again apply the chain rule, but this time to e^{x^2} , and get

$$c'(x) = \cos(e^{x^2}) [e^{x^2} \cdot 2x].$$

The chain rule now adds substantially to our ability to compute derivatives. Whether we are 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 is a composition, the chain rule is indispensable.

Example 2.5.9. Use known derivative rules, including the chain rule, as needed to answer each of the following questions.

- Find an equation for the tangent line to the curve $y = \sqrt{e^x + 3}$ at the point where $x = 0$.
- If $s(t) = \frac{1}{(t^2 + 1)^3}$ 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?
- 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 of 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.
- 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

Table 2.5.10: Data for functions f and g .

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 $D'(-1)$.

Hint.

- a. Let $f(x) = \sqrt{e^x + 3}$. Find $f'(x)$ and $f'(0)$.
- b. Recall that $s'(t)$ tells us the instantaneous velocity at time t .
- c. Note that the units of dP/dh are inches of mercury per foot.
- d. Since $C(x) = f(g(x))$, it follows $C'(x) = f'(g(x))g'(x)$. What is $C'(2)$?

Answer.

- a. $y - 2 = \frac{1}{4}(x - 0)$.
- b. $P'(1000) = 30e^{-0.0323}(-0.0000323) \approx -0.000938$ inches of mercury per foot.
- c. $C'(2) = -10$; $D'(-1) = -20$.

Solution.

- a. Let $f(x) = \sqrt{e^x + 3}$. By the chain rule $f'(x) = \frac{e^x}{2\sqrt{e^x + 3}}$, and thus $f'(0) = \frac{1}{4}$. Note further that $f(0) = \sqrt{1 + 3} = 2$. The tangent line is therefore the line through $(0, 2)$ with slope $1/4$, which is

$$y - 2 = \frac{1}{4}(x - 0).$$

- b. Observe that $s(t) = (t^2 + 1)^{-3}$, and thus by the chain rule, $s'(t) = -3(t^2 + 1)^{-4}(2t)$. We therefore see that $s'(1) = -\frac{6}{16} = -\frac{3}{8}$ inches per second, so the particle is moving left at the instant $t = 1$.
- c. First, $P'(h) = \frac{dP}{dh} = 30e^{-0.0323h}(-0.0000323)$. Therefore, $P'(1000) = 30e^{-0.0323}(-0.0000323) \approx -0.000938$ inches of mercury per foot. This tells us that the barometric pressure is dropping very slightly for each additional foot of elevation gain.
- d. Since $C(x) = f(g(x))$, it follows $C'(x) = f'(g(x))g'(x)$. Therefore, $C'(2) = f'(g(2))g'(2)$. From the given table, $g(2) = -1$, so applying this result and using additional given information,

$$C'(2) = f'(-1)g'(2) = (-5)(2) = -10.$$

For $D(x) = f(f(x))$, the chain rule tells us that $D'(x) = f'(f(x))f'(x)$, so $D'(-1) = f'(f(-1))f'(-1)$. Using the given table, it follows

$$D'(-1) = f'(2)f'(-1) = (4)(-5) = -20.$$

2.5.3 The composite version of basic function rules

As we gain more experience with differentiation, we will become more comfortable in simply writing down the derivative without taking multiple steps. This is particularly simple when the inner function is linear, since the derivative of a linear function is a constant.

Example 2.5.11. Use the chain rule to differentiate each of the following composite functions whose inside function is linear:

$$\frac{d}{dx} [(5x + 7)^{10}] = 10(5x + 7)^9 \cdot 5,$$

$$\frac{d}{dx} [\tan(17x)] = 17 \sec^2(17x), \text{ and}$$

$$\frac{d}{dx} [e^{-3x}] = -3e^{-3x}.$$

More generally, following is an excellent exercise for getting comfortable with the derivative rules. Write down a list of all the basic functions whose derivatives we know, and list the derivatives. Then write a composite function with the inner function being an unknown function $u(x)$ and the outer function being a basic function. Finally, write the chain rule for the composite function. The following example illustrates this for two different functions.

Example 2.5.12. To determine

$$\frac{d}{dx} [\sin(u(x))],$$

where u is a differentiable function of x , we use the chain rule with the sine function as the outer function. Applying the chain rule, we find that

$$\frac{d}{dx} [\sin(u(x))] = \cos(u(x)) \cdot u'(x).$$

This rule is analogous to the basic derivative rule that $\frac{d}{dx} [\sin(x)] = \cos(x)$.

Similarly, since $\frac{d}{dx} [a^x] = a^x \ln(a)$, it follows by the chain rule that

$$\frac{d}{dx} [a^{u(x)}] = a^{u(x)} \ln(a) \cdot u'(x).$$

This rule is analogous to the basic derivative rule that $\frac{d}{dx} [a^x] = a^x \ln(a)$.

2.5.4 Summary

- 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 $x \rightarrow \sin(x) \rightarrow 2^{\sin(x)}$.
- Given a composite function $C(x) = f(g(x))$ where f and g are differentiable functions, the chain rule tells us that

$$C'(x) = f'(g(x))g'(x).$$

2.5.5 Exercises

1. (Mixing rules: chain, product, sum) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

2. (Mixing rules: chain and product) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Using the chain rule repeatedly) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Derivative involving arbitrary constants a and b) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Chain rule with graphs) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Chain rule with function values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (A product involving a composite function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. 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?

Answer.

- a. $h'(\frac{\pi}{4}) = \frac{3}{2\sqrt{2}}$.
- b. $r'(0.25) = \cos(0.25^3) \cdot 3(0.25)^2 \approx 0.1875 > h'(0.25) = 3\sin^2(0.25) \cdot \cos(0.25) \approx 0.1779$; r is changing more rapidly.
- c. $h'(x)$ is periodic; $r'(x)$ is not.

Solution.

- a. Since $h(x) = f(g(x))$, by the chain rule we know that $h'(x) = f'(g(x)) \cdot g'(x)$. Using the facts that $f(x) = x^3$ and $g(x) = \sin(x)$, it follows $f'(x) = 3x^2$ and $f'(g(x)) = 3\sin^2(x)$, and $g'(x) = \cos(x)$. Therefore,

$$h'(x) = 3\sin^2(x) \cdot \cos(x)$$

$$\text{and } h'(\frac{\pi}{4}) = 3 \left(\frac{1}{\sqrt{2}} \right)^2 \cdot \frac{1}{\sqrt{2}} = \frac{3}{2\sqrt{2}}.$$

- b. Note that $r'(x) = g'(f(x)) \cdot f'(x) = \cos(x^3) \cdot 3x^2$. Thus, we have $r'(0.25) = \cos(0.25^3) \cdot 3(0.25)^2 \approx 0.1875$, while using our work from (a), $h'(0.25) = 3\sin^2(0.25) \cdot \cos(0.25) \approx 0.1779$, so since $|h'(0.25)| < |r'(0.25)|$, we see that r is changing more rapidly.
- c. We have observed in (a) and (b) that $h'(x) = 3\sin^2(x) \cdot \cos(x)$ and $r'(x) = \cos(x^3) \cdot 3x^2$. Because both parts of the product in $h'(x)$ are periodic functions, $h'(x)$ is periodic (with period 2π). However, $r'(x)$ is not periodic, since $3x^2$ grows without bound as x increases.
9. 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. $b(x) = u^4(x)$

Answer.

- a. $p'(x) = e^{u(x)} \cdot u'(x)$.
- b. $q'(x) = u'(e^x) \cdot e^x$.
- c. $r'(x) = -\csc^2(u(x)) \cdot u'(x)$.
- d. $s'(x) = u'(\cot(x)) \cdot (-\csc^2(x))$.
- e. $a'(x) = u'(x^4) \cdot 4x^3$.
- f. $b'(x) = 4(u(x))^3 \cdot u'(x)$.

Solution.

- a. By the chain rule, $p'(x) = e^{u(x)} \cdot u'(x)$.
- b. By the chain rule, $q'(x) = u'(e^x) \cdot e^x$.
- c. By the chain rule, $r'(x) = -\csc^2(u(x)) \cdot u'(x)$.
- d. By the chain rule, $s'(x) = u'(\cot(x)) \cdot (-\csc^2(x))$.
- e. By the chain rule, $a'(x) = u'(x^4) \cdot 4x^3$.
- f. By the chain rule, $b'(x) = 4(u(x))^3 \cdot u'(x)$.

10. Let functions p and q be the piecewise linear functions given by their respective graphs in Figure 2.5.13. Use the graphs to answer the following questions.

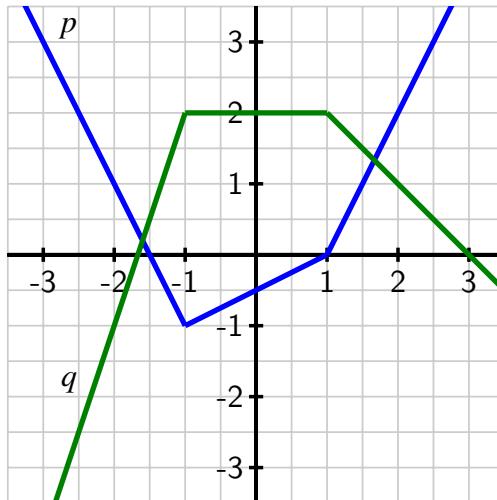


Figure 2.5.13: The graphs of p (in blue) and q (in green).

- Let $C(x) = p(q(x))$. Determine $C'(0)$ and $C'(3)$.
- Find a value of x for which $C'(x)$ does not exist. Explain your thinking.
- Let $Y(x) = q(q(x))$ and $Z(x) = q(p(x))$. Determine $Y'(-2)$ and $Z'(0)$.

Answer.

- $C'(0) = 0$ and $C'(3) = -\frac{1}{2}$.
- Consider $C'(1)$. By the chain rule, we'd expect that $C'(1) = p'(q(1)) \cdot q'(1)$, but we know that $q'(1)$ does not exist since q has a corner point at $x = 1$. This means that $C'(1)$ does not exist either.
- Since $Y(x) = q(q(x))$, the chain rule implies that $Y'(x) = q'(q(x)) \cdot q'(x)$, and thus $Y'(-2) = q'(q(-2)) \cdot q'(-2) = q'(-1) \cdot q'(-2)$. But $q'(-1)$ does not exist, so $Y'(-2)$ also fails to exist. Using $Z(x) = q(p(x))$ and the chain rule, we have $Z'(x) = q'(p(x)) \cdot p'(x)$. Therefore $Z'(0) = q'(p(0)) \cdot p'(0) = q'(-0.5) \cdot p'(0) = 0 \cdot 0.5 = 0$.

Solution.

- By the chain rule, $C'(x) = p'(q(x)) \cdot q'(x)$, so $C'(0) = p'(q(0)) \cdot q'(0)$ and $C'(3) = p'(q(3)) \cdot q'(3)$. From the graph, we see that $q(0) = 2$, and thus $C'(0) = p'(2) \cdot q'(0)$. Reading the slopes of the lines from the graph, it follows $C'(0) = 2 \cdot 0 = 0$. Similarly, $q(3) = 0$, so $C'(3) = p'(0) \cdot q'(3) = \frac{1}{2} \cdot (-1) = -\frac{1}{2}$.
- Consider $C'(1)$. By the chain rule, we'd expect that $C'(1) = p'(q(1)) \cdot q'(1)$, but we know that $q'(1)$ does not exist since q has a corner point at $x = 1$. This means that $C'(1)$ does not exist either.
- Since $Y(x) = q(q(x))$, the chain rule implies that $Y'(x) = q'(q(x)) \cdot q'(x)$, and thus $Y'(-2) = q'(q(-2)) \cdot q'(-2) = q'(-1) \cdot q'(-2)$. But $q'(-1)$ does not exist, so $Y'(-2)$ also fails to exist. Using $Z(x) = q(p(x))$ and the chain rule, we have $Z'(x) = q'(p(x)) \cdot p'(x)$. Therefore $Z'(0) = q'(p(0)) \cdot p'(0) = q'(-0.5) \cdot p'(0) = 0 \cdot 0.5 = 0$.

11. 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

$$V = \frac{\pi}{3}h^2(12 - h).$$

- 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?
- 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 $t = 2$?
- 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 $t = 2$?
- What are the main differences between the rates found in (a) and (c)? Include a discussion of the relevant units.

Answer.

- $\frac{dV}{dh} = \pi(8h - h^2)$, cubic feet per foot.

- b. $\frac{dV}{dt} = \frac{\pi}{3} [12 \cdot 2(\sin(\pi t) + 1) \cdot \pi \cos(\pi t) - 3(\sin(\pi t) + 1)^2 \cdot \pi \cos(\pi t)]$ cubic feet per hour.
- c. $\frac{dV}{dt} \Big|_{t=0} = 7\pi^2$ cubic feet per hour.
- d. In (a) we are determining the instantaneous rate at which the volume changes as we increase the height of the water in the tank, while in (c) we are finding the instantaneous rate at which volume changes as we increase time.

Solution.

- a. We are asked to find $\frac{dV}{dh}$ (the derivative of V with respect to the independent variable h) at the instant $h = 1$. If we first expand V to write

$$V = \frac{\pi}{3} (12h^2 - h^3)$$

we find that

$$\frac{dV}{dh} = \frac{\pi}{3} (24h - 3h^2) = \pi (8h - h^2),$$

with units “cubic feet per foot”.

- b. Now we want to find $\frac{dV}{dt}$, the derivative of V with respect to the independent variable t . Here we can write V as a function of t by writing $V(t) = \frac{\pi}{3} (12(h(t))^2 - (h(t))^3)$, where $h(t) = \sin(\pi t) + 1$. By the chain rule, it follows that

$$\frac{dV}{dt} = \frac{\pi}{3} (12 \cdot 2h(t) \cdot h'(t) - 3(h(t))^2 \cdot h'(t))$$

and hence using $h(t) = \sin(\pi t) + 1$ and $h'(t) = \pi \cos(\pi t)$,

$$\frac{dV}{dt} = \frac{\pi}{3} [12 \cdot 2(\sin(\pi t) + 1) \cdot \pi \cos(\pi t) - 3(\sin(\pi t) + 1)^2 \cdot \pi \cos(\pi t)]$$

with units “cubic feet per hour”.

- c. Now we want to find $\frac{dV}{dt}$ when $t = 0$. Using the result from part (b), we have

$$\frac{dV}{dt} \Big|_{t=0} = \frac{\pi}{3} [12 \cdot 2(\sin(0) + 1) \cdot \pi \cos(0) - 3(\sin(0) + 1)^2 \cdot \pi \cos(0)] = 7\pi^2$$

cubic feet per hour.

- d. In (a) we are determining the instantaneous rate at which the volume changes as we increase the height of the water in the tank, while in (c) we are finding the instantaneous rate at which volume changes as we increase time. The first calculation is a direct one, while the second is more indirect, as it involves the composition of V with h as a function of time. In the latter case, both h and V are functions of time, which allows us to think not only of how V is changing as a function of h , but also as a function of t .

2.6 Derivatives of Inverse Functions

Motivating 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 ?
-

Much of mathematics centers on the notion of function. Indeed, throughout our study of calculus, we are investigating the behavior of functions, 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? And if we know how fast a particular process is changing, can we determine how fast the inverse process is changing?

One of the most important functions in all of mathematics is the natural exponential function $f(x) = e^x$. Its inverse, the natural logarithm, $g(x) = \ln(x)$, is similarly important. One of our goals in this section is to learn how to differentiate the logarithm function. First, we review some of the basic concepts surrounding functions and their inverses.

Example 2.6.1. 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?

2.6.1 Basic facts about inverse functions

A function $f : A \rightarrow B$ 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 $g : B \rightarrow A$ 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 . The inverse function undoes the work of f . Indeed, if $y = f(x)$, then

$$f^{-1}(y) = f^{-1}(f(x)) = x.$$

Thus, the equations $y = f(x)$ and $x = f^{-1}(y)$ say the 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.

Note 2.6.2. For a function $f : A \rightarrow B$,

- f has an inverse if and only if f is one-to-one¹ and onto²;
- 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 fact reveals a special relationship between the graphs of f and f^{-1} . If a point (x, y) that lies on the graph of $y = f(x)$, 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 each other across the line $y = x$, because this reflection is precisely the geometric action that swaps the coordinates in an ordered pair. In Figure 2.6.3, we see this illustrated by 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 $y = x$.

¹A function f is *one-to-one* provided that no two distinct inputs lead to the same output.

²A 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.

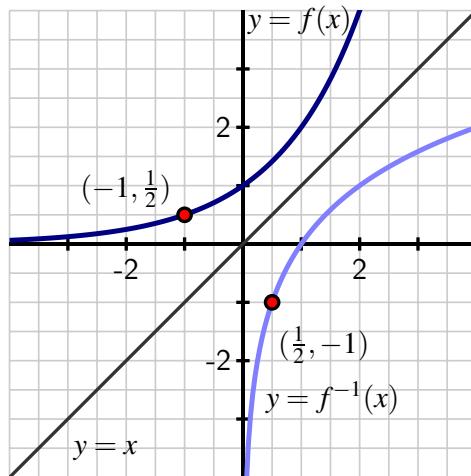


Figure 2.6.3: A graph of a function $y = f(x)$ along with its inverse, $y = f^{-1}(x)$.

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, namely the natural logarithm, $x = f^{-1}(y) = \ln(y)$. Thus, 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 .

2.6.2 The derivative of the natural logarithm function

In what follows, we find 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, 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 equation and observe that

$$\frac{d}{dx} [e^{g(x)}] = \frac{d}{dx}[x].$$

The righthand side is simply 1; by applying the chain rule to the left side, we find that

$$e^{g(x)} g'(x) = 1.$$

Next we solve for $g'(x)$, to get

$$g'(x) = \frac{1}{e^{g(x)}}.$$

Finally, we recall that $g(x) = \ln(x)$, so $e^{g(x)} = e^{\ln(x)} = x$, 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 basic derivative rules. Note that this rule applies only to positive values of x , as these are the only values for which $\ln(x)$ is defined.

Also notice 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, $f'(t) = \frac{1}{t^2+1} \cdot 2t$.

There are interesting connections between the graphs of $f(x) = e^x$ and $f^{-1}(x) = \ln(x)$.

In Figure 2.6.4, 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 $m_B = 5$.

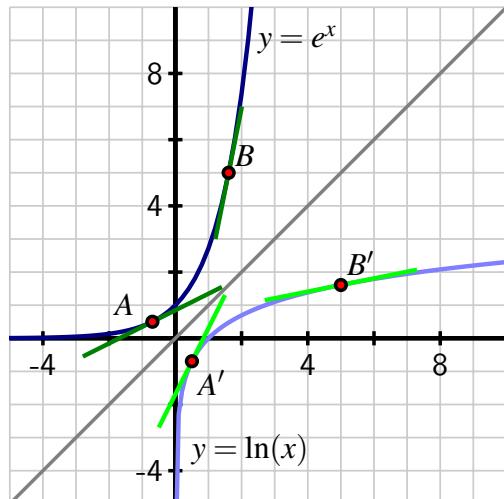


Figure 2.6.4: A graph of the function $y = e^x$ along with its inverse, $y = \ln(x)$, where both functions are viewed using the input variable x .

At the corresponding points A' and B' on the graph of the natural logarithm function (which come from reflecting A and B 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, at $A' = (0.5, \ln(0.5))$, 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. This is due to the reflection across $y = x$. It changes the roles of x and y , thus reversing the rise and run, so the slope of the inverse function at the reflected point is the reciprocal of the slope of the

original function.

Example 2.6.5. For each function given below, find its derivative.

- | | |
|------------------------------------|--------------------------|
| a. $h(x) = x^2 \ln(x)$ | d. $z(x) = \tan(\ln(x))$ |
| b. $p(t) = \frac{\ln(t)}{e^t + 1}$ | |
| c. $s(y) = \ln(\cos(y) + 2)$ | e. $m(z) = \ln(\ln(z))$ |

Hint.

- a. Is h a product, quotient, or composition of basic functions?
- b. Is p a product, quotient, or composition of basic functions?
- c. Is s a product, quotient, or composition of basic functions?
- d. Is z a product, quotient, or composition of basic functions?
- e. Is m a product, quotient, or composition of basic functions?

Answer.

- a. $h'(x) = x + 2x \ln(x).$
- b. $p'(t) = \frac{(e^t + 1)\frac{1}{t} - \ln(t) \cdot e^t}{(e^t + 1)^2}.$
- c. $s'(y) = \frac{1}{\cos(y) + 2} \cdot (-\sin(y)).$
- d. $z'(x) = \sec^2(\ln(x)) \cdot \frac{1}{x}.$
- e. $m'(z) = \frac{1}{\ln(z)} \cdot \frac{1}{z}.$

Solution.

- a. By the product rule,

$$h'(x) = x^2 \cdot \frac{1}{x} + \ln(x) \cdot 2x = x + 2x \ln(x).$$

- b. By the quotient rule,

$$p'(t) = \frac{(e^t + 1)\frac{1}{t} - \ln(t) \cdot e^t}{(e^t + 1)^2}.$$

- c. The chain rule tells us that

$$s'(y) = \frac{1}{\cos(y) + 2} \cdot (-\sin(y)).$$

d. Again using the chain rule,

$$z'(x) = \sec^2(\ln(x)) \cdot \frac{1}{x}.$$

e. Noting that m is composite with the natural logarithm function serving as both the inner and outer function, we find that

$$m'(z) = \frac{1}{\ln(z)} \cdot \frac{1}{z}.$$

2.6.3 Inverse trigonometric functions and their derivatives

Trigonometric functions are periodic, so they fail to be one-to-one, and thus do not have inverse functions. However, we can restrict the domain of each trigonometric function so that it is one-to-one on that domain.

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. Thus, the function $f(x) = \sin(x)$ with $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and codomain $[-1, 1]$ has an inverse function f^{-1} such that

$$f^{-1} : [-1, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}].$$

We call f^{-1} the *arcsine* (or inverse sine) function and write $f^{-1}(y) = \arcsin(y)$. It is especially important to remember that

$$y = \sin(x) \text{ and } x = \arcsin(y)$$

say the same thing. “The arcsine of y ” means “the angle whose sine is y .” For example, $\arcsin(\frac{1}{2}) = \frac{\pi}{6}$ means that $\frac{\pi}{6}$ is the angle whose sine is $\frac{1}{2}$, which is equivalent to writing $\sin(\frac{\pi}{6}) = \frac{1}{2}$.

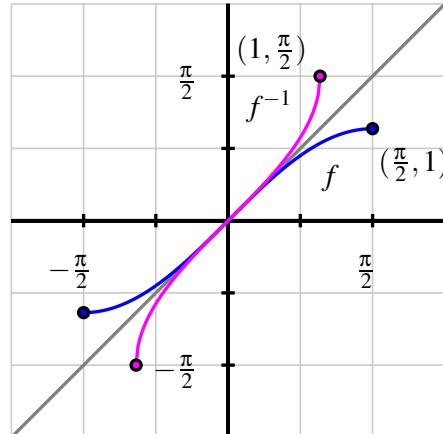


Figure 2.6.6: A graph of $f(x) = \sin(x)$ (in blue), restricted to the domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$, along with its inverse, $f^{-1}(x) = \arcsin(x)$ (in magenta).

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 angle whose sine is x , it is equivalent to write

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

Differentiating both sides of the previous equation, we have

$$\frac{d}{dx}[\sin(h(x))] = \frac{d}{dx}[x].$$

The righthand side is simply 1, and by applying 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.

Let’s say that $\theta = \arcsin(x)$, so that θ is the angle whose sine is x . 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 2.6.7. The horizontal leg must be $\sqrt{1 - x^2}$, by the Pythagorean Theorem.

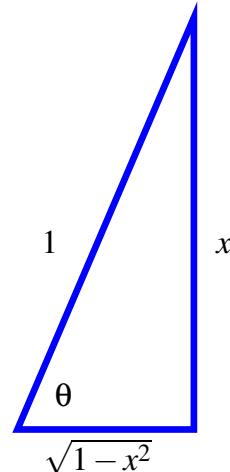


Figure 2.6.7: The right triangle that corresponds to the angle $\theta = \arcsin(x)$.

Now, because $\theta = \arcsin(x)$, the expression for $\cos(\arcsin(x))$ is equivalent to $\cos(\theta)$. From the figure, $\cos(\arcsin(x)) = \cos(\theta) = \sqrt{1 - x^2}$.

Substituting this expression into our formula, $h'(x) = \frac{1}{\cos(\arcsin(x))}$, we have now shown that

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

Inverse sine

For all real numbers x such that $-1 < x < 1$,

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

Example 2.6.8. The following prompts in this example will lead you to develop the derivative of the inverse tangent function.

- Let $r(x) = \arctan(x)$. Use the relationship between the arctangent and tangent functions to rewrite this equation using only the tangent function.
- 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 $r(x)$.
- Recall that $r(x) = \arctan(x)$. Update your expression for $r'(x)$ so that it only involves trigonometric functions and the independent variable x .
- Introduce a right triangle with angle θ so that $\theta = \arctan(x)$. What are the three sides of the triangle?
- In terms of only x and 1, what is the value of $\cos(\arctan(x))$?
- Use the results of your work above to find an expression involving only 1 and x for $r'(x)$.

Hint.

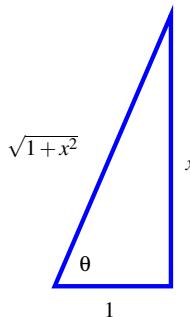
- Recall that for any function and its inverse, writing $y = f^{-1}(x)$ is equivalent to writing $f(y) = x$.
- Apply the chain rule to differentiate $\tan(r(x))$.
- This question is only asking you to substitute the expression for $r(x)$ into what you found in (b).
- Let the vertical leg of the triangle be x . What must the horizontal leg be?
- Recall that the cosine of an angle is the length of the adjacent leg over the length of the hypotenuse.
- Note that $\cos^2(\alpha) = (\cos(\alpha))^2$.

Answer.

- $\tan(r(x)) = x$.
- $r'(x) = \cos^2(r(x))$.

c. $r'(x) = \cos^2(\arctan(x))$.

d. With $\theta = \arctan(x)$,



e. $\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}}$.

f. $r'(x) = \frac{1}{1+x^2}$.

Solution.

a. Since $r(x) = \arctan(x)$, it is equivalent to write $\tan(r(x)) = x$.

b. Differentiating, we have $\frac{d}{dx}[\tan(r(x))] = \frac{d}{dx}[x]$, so

$$\sec^2(r(x))r'(x) = 1,$$

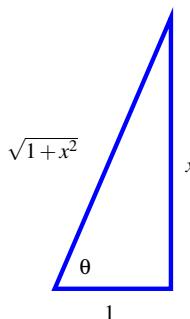
and thus

$$r'(x) = \frac{1}{\sec^2(r(x))} = \cos^2(r(x)),$$

since $\frac{1}{\sec(\alpha)} = \cos(\alpha)$.

c. Since $r(x) = \arctan(x)$, we now have that $r'(x) = \cos^2(\arctan(x))$.

d. Letting $\theta = \arctan(x)$, it follows that we can view θ as an angle in a right triangle with legs 1 and x (so that $\tan(\theta) = \frac{x}{1}$, and thus by the Pythagorean Theorem, the triangle's hypotenuse is $\sqrt{1+x^2}$, as shown below.



e. To evaluate $\cos(\arctan(x))$, we use the right triangle developed in (d) and observe that

$$\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}}.$$

f. Finally, we recall that we know $r'(x) = \cos^2(\arctan(x)) = (\cos(\arctan(x)))^2$. Having established that $\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}}$, we now have that

$$r'(x) = \frac{1}{1+x^2}.$$

While derivatives for other inverse trigonometric functions can be established similarly, for now we limit ourselves to the arcsine and arctangent functions.

Example 2.6.9. 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. $g(w) = \arctan\left(\frac{\ln(w)}{1+w^2}\right)$

Hint.

- a. Use the sum rule followed by the product rule on each term in the sum.
- b. Use the chain rule first. What rule is needed to differentiate $t \arcsin(t)$?
- c. Note that the chain rule is needed to differentiate $\arcsin(5z)$.
- d. You can use right triangle trigonometry to simplify the function first.
- e. Use the chain rule twice.
- f. Think about the overall structure of this function.

Answer.

- a. $f'(x) = \left[x^3 \cdot \frac{1}{1+x^2} + \arctan(x) \cdot 3x^2\right] + \left[e^x \cdot \frac{1}{x} + \ln(x) \cdot e^x\right].$
- b. $p'(t) = 2^{t \arcsin(t)} \ln(2) \left[t \cdot \frac{1}{\sqrt{1-t^2}} + \arcsin(t) \cdot 1\right].$
- c. $h'(z) = 27(\arcsin(5z) + \arctan(4 - z))^{26} \left[\frac{1}{1-(5z)^2} \cdot 5 + \frac{1}{1+(4-z)^2} \cdot (-1)\right].$

d. $s'(y) = -\frac{1}{y^2}$.

e. $m'(v) = \frac{1}{\sin^2(v)+1} \cdot [2 \sin(v) \cos(v)]$.

f. $g'(w) = \frac{1}{1+\left(\frac{\ln(w)}{1+w^2}\right)^2} \cdot \left[\frac{(1+w^2)\frac{1}{w} - \ln(w) \cdot 2w}{(1+w^2)^2} \right]$

Solution.

a. By the sum rule followed by two applications of the product rule,

$$\begin{aligned} f'(x) &= \frac{d}{dx}[x^3 \arctan(x)] + \frac{d}{dx}[e^x \ln(x)] \\ &= \left[x^3 \cdot \frac{1}{1+x^2} + \arctan(x) \cdot 3x^2 \right] + \left[e^x \cdot \frac{1}{x} + \ln(x) \cdot e^x \right] \end{aligned}$$

b. Using the chain rule followed by the product rule,

$$\begin{aligned} p'(t) &= 2^{t \arcsin(t)} \ln(2) \frac{d}{dt}[t \arcsin(t)] \\ &= 2^{t \arcsin(t)} \ln(2) [t \cdot \frac{1}{\sqrt{1-t^2}} + \arcsin(t) \cdot 1] \end{aligned}$$

c. Applying the chain rule twice,

$$\begin{aligned} h'(z) &= 27(\arcsin(5z) + \arctan(4-z))^{26} \frac{d}{dz}[\arcsin(5z) + \arctan(4-z)] \\ &= 27(\arcsin(5z) + \arctan(4-z))^{26} \left[\frac{1}{1-(5z)^2} \cdot \frac{d}{dz}[5z] + \frac{1}{1+(4-z)^2} \cdot \frac{d}{dz}[4-z] \right] \\ &= 27(\arcsin(5z) + \arctan(4-z))^{26} \left[\frac{1}{1-(5z)^2} \cdot 5 + \frac{1}{1+(4-z)^2} \cdot (-1) \right] \end{aligned}$$

d. Using right triangle trigonometry, it is straightforward to show that $\cot(\arctan(y)) = \frac{1}{y}$.
As such, $s'(y) = -\frac{1}{y^2}$.

e. By two applications of the chain rule (noting particularly that $\sin^2(v) = (\sin(v))^2$, we have

$$\begin{aligned} m'(v) &= \frac{1}{\sin^2(v)+1} \cdot \frac{d}{dv}[\sin^2(v)+1] \\ &= \frac{1}{\sin^2(v)+1} \cdot [2 \sin(v) \cos(v)] \end{aligned}$$

f. By the chain rule followed by the quotient rule,

$$\begin{aligned} g'(w) &= \frac{1}{1+\left(\frac{\ln(w)}{1+w^2}\right)^2} \cdot \frac{d}{dw}\left[\frac{\ln(w)}{1+w^2}\right] \\ &= \frac{1}{1+\left(\frac{\ln(w)}{1+w^2}\right)^2} \cdot \left[\frac{(1+w^2)\frac{1}{w} - \ln(w) \cdot 2w}{(1+w^2)^2} \right] \end{aligned}$$

2.6.4 The link between the derivative of a function and the derivative of its inverse

In Figure 2.6.4, we saw an interesting relationship between the slopes of tangent lines to the natural exponential and natural logarithm functions at points reflected across the line $y = x$. In particular, we observed that at the point $(\ln(2), 2)$ on the graph of $f(x) = e^x$, the slope of the tangent line 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 is $(f^{-1})'(2) = \frac{1}{2}$, which is the reciprocal of $f'(\ln(2))$.

That the two corresponding tangent lines have reciprocal slopes is not a coincidence. If f and g are differentiable inverse functions, then $y = f(x)$ if and only if $x = g(y)$, then $f(g(x)) = x$ for every x in the domain of f^{-1} . Differentiating both sides of this equation, 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 $(g(x), f(g(x))) = (g(x), x)$.

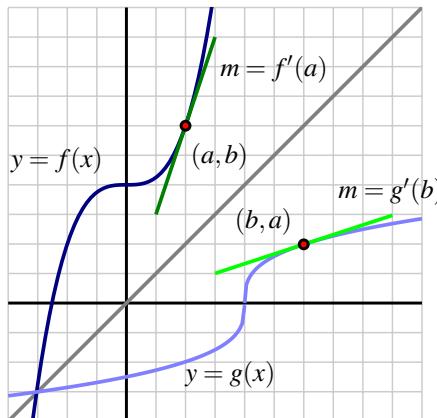


Figure 2.6.10: A graph of function $y = f(x)$ along with its inverse, $y = g(x) = f^{-1}(x)$. Observe that the slopes of the two tangent lines are reciprocals of one another.

To see this more clearly, consider the graph of the function $y = f(x)$ shown in Figure 2.6.10, 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 ; because $f(a) = b$ and $g(b) = a$. Now, applying the rule that

$g'(x) = 1/f'(g(x))$ 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 $y = x$.

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 $g'(x) = 1/f'(g(x))$.

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

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{e^{\ln(x)}} = \frac{1}{x}.$$

2.6.5 Summary

- For all positive real numbers x , $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$.
- For all real numbers x such that $-1 < x < 1$, $\frac{d}{dx}[\arcsin(x)] = \frac{1}{\sqrt{1-x^2}}$. 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' , $g'(x) = \frac{1}{f'(g(x))}$.

2.6.6 Exercises

1. (Composite function involving logarithms and polynomials) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Composite function involving trigonometric functions and logarithms) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Product involving $\arcsin(w)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

4. (Derivative involving $\arctan(x)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Composite function from a graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Composite function involving an inverse trigonometric function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Mixing rules: product, chain, and inverse trig) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. (Mixing rules: product and inverse trig) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
9. Determine the derivative of each of the following functions. Use proper notation and clearly identify the derivative rules you use.

- $f(x) = \ln(2 \arctan(x) + 3 \arcsin(x) + 5)$
- $r(z) = \arctan(\ln(\arcsin(z)))$
- $q(t) = \arctan^2(3t) \arcsin^4(7t)$
- $g(v) = \ln\left(\frac{\arctan(v)}{\arcsin(v)+v^2}\right)$

Answer.

- $f'(x) = \frac{1}{2 \arctan(x) + 3 \arcsin(x) + 5} \cdot \left(\frac{2}{1+x^2} + \frac{3}{\sqrt{1-x^2}} \right).$
- $r'(z) = \frac{1}{1 + (\ln(\arcsin(z)))^2} \cdot \left(\frac{1}{\arcsin(z)} \right) \cdot \frac{1}{\sqrt{1-z^2}}.$
- $q'(t) = \arctan^2(3t) \cdot \left[4 \arcsin^3(7t) \left(\frac{7}{\sqrt{1+(7t)^2}} \right) \right] + \arcsin^4(7t) \cdot \left[2 \arctan(3t) \left(\frac{3}{1+(3t)^2} \right) \right].$
- $g'(v) = \frac{1}{\arcsin(v)+v^2} \cdot \frac{(\arcsin(v)+v^2) \cdot \frac{1}{1+v^2} - \arctan(v) \cdot \left(\frac{1}{\sqrt{1-v^2}} + 2v \right)}{(\arcsin(v)+v^2)^2}$

Solution.

- a. By the chain rule,

$$f'(x) = \frac{1}{2 \arctan(x) + 3 \arcsin(x) + 5} \cdot \frac{d}{dx} (2 \arctan(x) + 3 \arcsin(x) + 5),$$

and by the sum and constant multiple rules, it follows that

$$f'(x) = \frac{1}{2 \arctan(x) + 3 \arcsin(x) + 5} \cdot \left(\frac{2}{1+x^2} + \frac{3}{\sqrt{1-x^2}} \right).$$

- b. By the chain rule applied to the outermost function,

$$r'(z) = \frac{1}{1 + (\ln(\arcsin(z)))^2} \cdot \frac{d}{dz} (\ln(\arcsin(z))).$$

Applying the chain rule again to evaluate $\frac{d}{dz} (\ln(\arcsin(z)))$,

$$r'(z) = \frac{1}{1 + (\ln(\arcsin(z)))^2} \cdot \left(\frac{1}{\arcsin(z)} \right) \cdot \frac{d}{dz} (\arcsin(z)).$$

Finally, using the basic rule for $\frac{d}{dz}(\arcsin(z))$,

$$r'(z) = \frac{1}{1 + (\ln(\arcsin(z)))^2} \cdot \left(\frac{1}{\arcsin(z)} \right) \cdot \frac{1}{\sqrt{1 - z^2}}.$$

c. First, by the product rule,

$$q'(t) = \arctan^2(3t) \cdot \frac{d}{dt}(\arcsin^4(7t)) + \arcsin^4(7t) \cdot \frac{d}{dt}(\arctan^2(3t)).$$

Next, we recall that $\arcsin^4(7t) = (\arcsin(7t))^4$ and $\arctan^2(3t) = (\arctan(3t))^2$, and apply the chain rule to find the derivative of each of these functions to get

$$q'(t) = \arctan^2(3t) \cdot \left[4 \arcsin^3(7t) \frac{d}{dt}(\arcsin(7t)) \right] + \arcsin^4(7t) \cdot \left[2 \arctan(3t) \frac{d}{dt}(\arctan(3t)) \right].$$

Finally, applying the chain rule one more time to evaluate the remaining derivatives,

$$q'(t) = \arctan^2(3t) \cdot \left[4 \arcsin^3(7t) \left(\frac{7}{\sqrt{1 + (7t)^2}} \right) \right] + \arcsin^4(7t) \cdot \left[2 \arctan(3t) \left(\frac{3}{1 + (3t)^2} \right) \right].$$

d. By the chain rule,

$$g'(v) = \frac{1}{\frac{\arctan(v)}{\arcsin(v) + v^2}} \cdot \frac{d}{dv} \left[\frac{\arctan(v)}{\arcsin(v) + v^2} \right]$$

Thus, by the quotient rule,

$$g'(v) = \frac{1}{\frac{\arctan(v)}{\arcsin(v) + v^2}} \cdot \frac{(\arcsin(v) + v^2) \cdot \frac{1}{1+v^2} - \arctan(v) \cdot \left(\frac{1}{\sqrt{1-v^2}} + 2v \right)}{(\arcsin(v) + v^2)^2}$$

10. Consider the graph of $y = f(x)$ provided in Figure 2.6.11 and use it to answer the following questions.

- Use the provided graph to estimate the value of $f'(1)$.
- 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 .
- Based on your work in (a), what is the value of $(f^{-1})'(-1)$? Why?

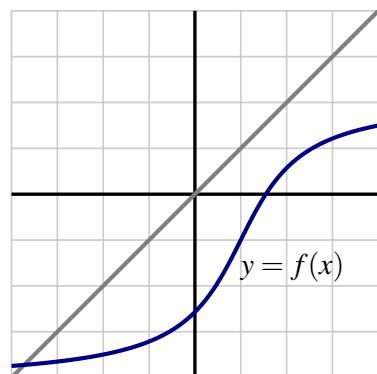
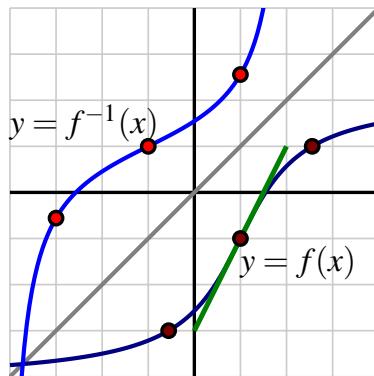


Figure 2.6.11: A function $y = f(x)$

Answer.

a. $f'(1) \approx 2$.

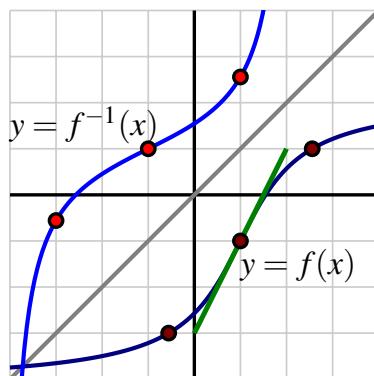
b.



c. $f'(-1) \approx 1/2$.

Solution.

- a. By plotting the tangent line as shown in the figure below, it appears that $f'(1) \approx 2$.
- b. One way to estimate the graph of $y = f^{-1}(x)$ is to plot several (x, y) points on the graph of $y = f(x)$ and then plot the corresponding points (y, x) with the coordinates reversed. Doing so and recalling that the graph of $y = f^{-1}(x)$ is the reflection of the graph of $y = f(x)$ across the line $y = x$, we see the following figure.



- c. Because the original function passes through the point $(1, -1)$, the inverse function passes through $(-1, 1)$. Moreover, since the tangent line to the original function at

$(1, -1)$ has slope $f'(1) \approx 2$, and the tangent line to the inverse function at $(-1, 1)$ corresponds to the reflection of the original tangent line across the line $y = x$, it follows that $f'(-1) \approx 1/2$. This can also be deduced from the “derivative of an inverse function” rule, whereby $f'(-1) = \frac{1}{f'(1)}$.

11. Let $f(x) = \frac{1}{4}x^3 + 4$.
- Sketch a graph of $y = f(x)$ and explain why f is an invertible function.
 - Let g be the inverse of f and determine a formula for g .
 - Compute $f'(x)$, $g'(x)$, $f'(2)$, and $g'(6)$. What is the special relationship between $f'(2)$ and $g'(6)$? Why?

Answer.

- f passes the horizontal line test.
- $f^{-1}(x) = g(x) = \sqrt[3]{4x - 16}$.
- $f'(x) = \frac{3}{4}x^2$; $f'(2) = 3$. $g'(x) = \frac{1}{3}(4x - 16)^{-2/3} \cdot 4$; $g'(6) = \frac{1}{3}$. These two derivative values are reciprocals.

Solution.

- f is an invertible function because its graph passes the horizontal line test.
- To find a formula for $g = f^{-1}$, we take the equation $y = \frac{1}{4}x^3 + 4$ and solve for x in terms of y . Doing so, we find that $y - 4 = \frac{1}{4}x^3$, so $x^3 = 4(y - 4)$, and thus

$$x = \sqrt[3]{4(y - 4)} = f^{-1}(y).$$

Writing the inverse as a function of x , we have found that $f^{-1}(x) = g(x) = \sqrt[3]{4x - 16}$.

- First, $f'(x) = \frac{3}{4}x^2$, and thus $f'(2) = 3$. Next, $g'(x) = \frac{1}{3}(4x - 16)^{-2/3} \cdot 4$, so that $g'(6) = \frac{1}{3} \cdot 8^{-2/3} \cdot 4 = \frac{1}{3} \cdot \frac{1}{4} \cdot 4 = \frac{1}{3}$. What is special about these two derivative values is that they happen to be reciprocals. Noting that $f(2) = 3$ and $g(3) = 2$, we know that $g'(3) = \frac{1}{f'(2)}$ by the rule for the derivative of an inverse function, and this matches our finding above that $g'(3) = \frac{1}{3}$.

12. Let $h(x) = x + \sin(x)$.

- Sketch a graph of $y = h(x)$ and explain why h must be invertible.
- Explain why it does not appear to be algebraically possible to determine a formula for h^{-1} .
- Observe that the point $(\frac{\pi}{2}, \frac{\pi}{2} + 1)$ lies on the graph of $y = h(x)$. Determine the value of $(h^{-1})'(\frac{\pi}{2} + 1)$.

Answer.

- h passes the horizontal line test.
- The equation $y = x + \sin(x)$ can't be solved for x in terms of y .
- $(h^{-1})'(\frac{\pi}{2} + 1) = 1$.

Solution.

- a. The function h is invertible because its graph passes the horizontal line test.
- b. It does not seem possible to solve the equation $y = x + \sin(x)$ for x in terms of y .
- c. By the rule for the derivative of an invertible function, we have $(h^{-1})'(\frac{\pi}{2} + 1) = \frac{1}{h'(\frac{\pi}{2})}$. Thus, we first compute $h'(x)$ and find that $h'(x) = 1 + \cos(x)$. Since $h'(\frac{\pi}{2}) = 1 + \cos(\frac{\pi}{2}) = 1$, it follows that $(h^{-1})'(\frac{\pi}{2} + 1) = \frac{1}{1} = 1$.

2.7 Derivatives of Functions Given Implicitly

Motivating 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?

In all of our studies with derivatives so far, we have worked with functions whose formula is given explicitly in terms of x . But there are many interesting curves whose equations involving x and y are impossible to solve for y in terms of x .

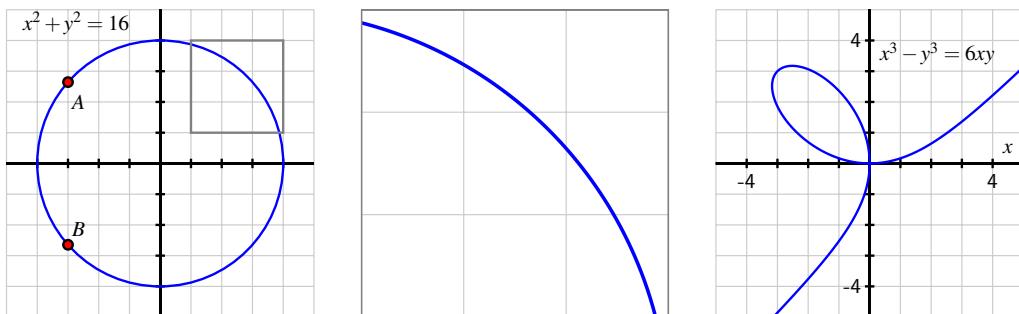


Figure 2.7.1: At left, the circle given by $x^2 + y^2 = 16$. In the middle, the portion of the circle $x^2 + y^2 = 16$ that has been highlighted in the box at left. And at right, the lemniscate given by $x^3 - y^3 = 6xy$.

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 2.7.1, 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 $y = f(x)$. But 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 2.7.1. 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 .

We say that the equation $x^2 + y^2 = 16$ defines y *implicitly* as a function of x . The graph

of the equation can be broken into pieces where each piece can be defined by an explicit function of x . For the circle, we could choose to take the top half as one function of x , namely $y = \sqrt{16 - x^2}$ and the bottom half as $y = -\sqrt{16 - x^2}$. The equation for the circle defines two *implicit functions* of x .

The righthand curve in Figure 2.7.1 is called a *lemniscate* and is just one of many fascinating possibilities for implicitly given curves.

How can we find an equation for $\frac{dy}{dx}$ without an explicit formula for y in terms of x ? The following example reminds us of some ways we can compute derivatives of functions in settings where the function's formula is not known.

Example 2.7.2. 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. $\frac{d}{dx}[x^2 + f(x)]$

d. $\frac{d}{dx}[f(x^2)]$

b. $\frac{d}{dx}[x^2 f(x)]$

c. $\frac{d}{dx}[c + x + f(x)^2]$

e. $\frac{d}{dx}[x f(x) + f(cx) + cf(x)]$

2.7.1 Implicit Differentiation

We begin our exploration of implicit differentiation with the example of the circle given by $x^2 + y^2 = 16$. How can we find a formula for $\frac{dy}{dx}$?

By viewing y as an *implicit* function of x , we 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.

So we view y as an unknown differentiable function of x and differentiate both sides of the equation with respect to x .

$$\frac{d}{dx}[x^2 + y^2] = \frac{d}{dx}[16].$$

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

$$\frac{d}{dx}[x^2] + \frac{d}{dx}[y^2] = 0.$$

Note carefully the different roles being played by x and y . Because x is the independent variable, $\frac{d}{dx}[x^2] = 2x$. But y is the dependent variable and y is an implicit function of x . Recall Example 2.7.2, where we computed $\frac{d}{dx}[f(x)^2]$. Computing $\frac{d}{dx}[y^2]$ is the same, and requires the chain rule, by which we find that $\frac{d}{dx}[y^2] = 2y^1 \frac{dy}{dx}$. We now have that

$$2x + 2y \frac{dy}{dx} = 0.$$

We solve this equation for $\frac{dy}{dx}$ by 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 . This makes sense because there are two corresponding points on the circle for each value of x between -4 and 4 , 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. The slope of the radius from the origin to the point (a, b) is $m_r = \frac{b}{a}$. The tangent line to the circle at (a, b) is perpendicular to the radius, and thus has slope $m_t = -\frac{a}{b}$, as shown in Figure 2.7.3. In particular, 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 $\frac{dy}{dx} = -\frac{x}{y}$.

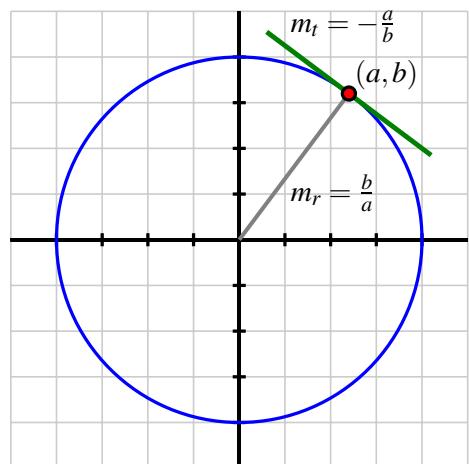


Figure 2.7.3: The circle given by $x^2 + y^2 = 16$ with point (a, b) on the circle and the tangent line at that point, with labeled slopes of the radial line, m_r , and tangent line, m_t .

Example 2.7.4. For the curve given implicitly by $x^3 + y^2 - 2xy = 2$, shown in Figure 2.7.5, find the slope of the tangent line at $(-1, 1)$.

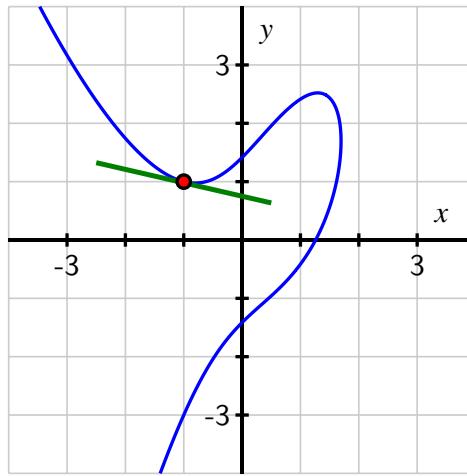


Figure 2.7.5: The curve $x^3 + y^2 - 2xy = 2$.

Solution. We begin by differentiating the curve's equation implicitly. Taking the derivative of each side with respect to x ,

$$\frac{d}{dx}[x^3 + y^2 - 2xy] = \frac{d}{dx}[2],$$

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

$$3x^2 + 2y \frac{dy}{dx} - [2x \frac{dy}{dx} + 2y] = 0.$$

We want to solve this equation for $\frac{dy}{dx}$. To do so, we first collect all of the terms involving $\frac{dy}{dx}$ on one side of the equation.

$$2y \frac{dy}{dx} - 2x \frac{dy}{dx} = 2y - 3x^2.$$

Then we factor the left side to isolate $\frac{dy}{dx}$.

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

Finally, we divide both sides by $(2y - 2x)$ and conclude that

$$\frac{dy}{dx} = \frac{2y - 3x^2}{2y - 2x}.$$

Note that 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 the coordinates into the formula for $\frac{dy}{dx}$, using the notation

$$\frac{dy}{dx}\bigg|_{(-1,1)} = \frac{2(1) - 3(-1)^2}{2(1) - 2(-1)} = -\frac{1}{4}.$$

This value matches our visual estimate of the slope of the tangent line shown in Figure 2.7.5.

Example 2.7.4 shows that it is possible when differentiating implicitly to have multiple terms involving $\frac{dy}{dx}$. We use addition and subtraction to collect all terms involving $\frac{dy}{dx}$ on one side of the equation, then factor to get a single term of $\frac{dy}{dx}$. Finally, we divide to solve for $\frac{dy}{dx}$.

We use the notation

$$\frac{dy}{dx}\bigg|_{(a,b)}$$

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.

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 2.7.6. Consider the curve defined by the equation $x = y^5 - 5y^3 + 4y$, whose graph is pictured in Figure 2.7.7.

- Explain why it is not possible to express y as an explicit function of x .
- Use implicit differentiation to find a formula for dy/dx .
- 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 $(0, 1)$.
- 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.

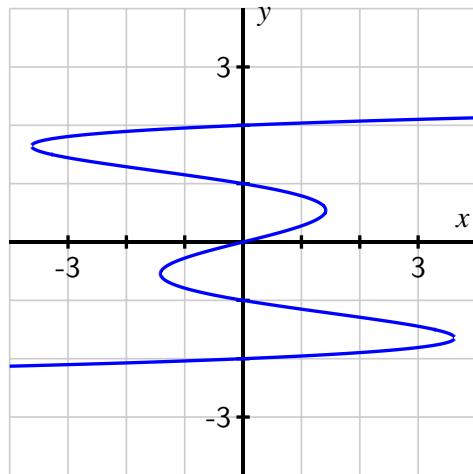


Figure 2.7.7: The curve $x = y^5 - 5y^3 + 4y$.

Hint.

- Does the graph pass the vertical line test?
- Note, for instance, that $\frac{d}{dx}[y^5] = 5y^4$.
- Remember the meaning of $\frac{dy}{dx}\big|_{(0,1)}$.
- What is the slope of a vertical line?

Answer.

- The graph of the curve fails the vertical line test.
- $\frac{dy}{dx} = \frac{1}{5y^4 - 15y^2 + 4}$.
- $y = -\frac{1}{6}x + 1$.
- $(1.418697, 0.543912), (-1.418697, -0.543912), (-3.63143, 1.64443)$, and $(3.63143, -1.64443)$.

Solution.

- Because the graph of the curve fails the vertical line test, y cannot be a function of x . This also confirms our intuition that there is not an algebraic means by which we can rearrange the equation $x = y^5 - 5y^3 + 4y$ to write y in terms of x .
- We differentiate implicitly, taking the derivative of each side with respect to x ,

$$\frac{d}{dx}[x] = \frac{d}{dx}[y^5 - 5y^3 + 4y],$$

and evaluate the elementary derivative on the left and use the sum rule on the right to find that

$$1 = \frac{d}{dx}[y^5] - \frac{d}{dx}[5y^3] + \frac{d}{dx}[4y].$$

By the chain and constant multiple rules, viewing y as a function of x , we now have

$$1 = 5y^4 \frac{dy}{dx} - 15y^2 \frac{dy}{dx} + 4 \frac{dy}{dx}.$$

Factoring,

$$1 = \frac{dy}{dx}(5y^4 - 15y^2 + 4),$$

and therefore

$$\frac{dy}{dx} = \frac{1}{5y^4 - 15y^2 + 4}.$$

- c. To find an equation of the line tangent to the graph of $x = y^5 - 5y^3 + 4y$ at the point $(0, 1)$, we only need the slope of the tangent line. Hence we compute

$$\left. \frac{dy}{dx} \right|_{(0,1)} = \frac{1}{5 \cdot 1^4 - 15 \cdot 1^2 + 4} = -\frac{1}{6}.$$

Therefore, the equation of the tangent line is

$$y - 1 = -\frac{1}{6}(x - 0)$$

or $y = -\frac{1}{6}x + 1$.

- d. Since a line is vertical whenever its slope is undefined, we seek all points (x, y) that make $\frac{dy}{dx}$ undefined. This will occur precisely when the denominator, $5y^4 - 15y^2 + 4$, is zero. Using a graphing utility or computer algebra system to solve the equation $5y^4 - 15y^2 + 4 = 0$, we find that this happens at the four approximate y -values $y \approx \pm 0.543912, \pm 1.64443$. For each such value, we use the original equation $x = y^5 - 5y^3 + 4y$ to find the x -value of the point. Doing so, we have established that there are four points at which the tangent line is vertical, and they are approximately $(1.418697, 0.543912)$, $(-1.418697, -0.543912)$, $(-3.63143, 1.64443)$, and $(3.63143, -1.64443)$.

It is natural to ask where the tangent line to a curve is vertical or horizontal. The slope of a horizontal tangent line must be zero, while the slope of a vertical tangent line is undefined.

Often the formula for $\frac{dy}{dx}$ is expressed as a quotient of functions of x and y , say

$$\frac{dy}{dx} = \frac{p(x, y)}{q(x, y)}.$$

The tangent line is horizontal precisely when the numerator is zero and the denominator is nonzero, making the slope of the tangent line zero. If we can solve the equation $p(x, y) = 0$ for either x and y in terms of the other, we can substitute that expression into the original equation for the curve. This gives an equation in a single variable, and if we can solve that

equation we can find the point(s) on the curve where $p(x, y) = 0$. At those points, the tangent line is horizontal.

Similarly, the tangent line is vertical whenever $q(x, y) = 0$ and $p(x, y) \neq 0$, making the slope undefined.

Example 2.7.8. Consider the curve defined by the equation $y(y^2 - 1)(y - 2) = x(x - 1)(x - 2)$, whose graph is pictured in Figure 2.7.9. 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.

- 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.)
- 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.)
- Find the equation of the tangent line to the curve at one of the points where $x = 1$.

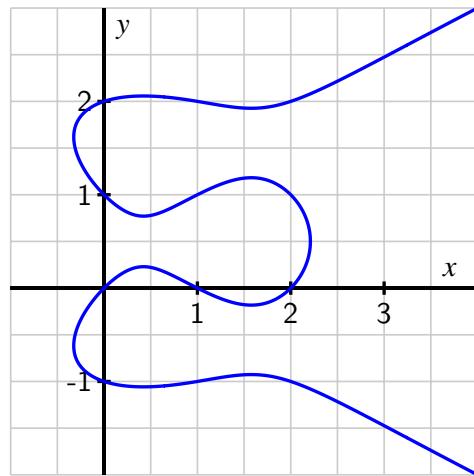


Figure 2.7.9: $y(y^2 - 1)(y - 2) = x(x - 1)(x - 2)$.

Hint.

- Note that the numerator of $\frac{dy}{dx}$ is a quadratic function of x .
- The denominator of $\frac{dy}{dx}$ is a cubic function of y .
- When $x = 1$, then y must satisfy the equation $y(y^2 - 1)(y - 2) = 0$.

Answer.

- Horizontal at $x \approx 0.42265$, thus $(0.42265, -1.05782); (0.42265, 0.229478); (0.42265, 0.770522); (0.42265, 2.05782)$. There are four more points where $x \approx 1.57735$.
- When $y = \frac{1}{2}, \frac{1 \pm \sqrt{5}}{2}$, so one point is $(2.21028, \frac{1}{2})$.
- $y - 1 = \frac{1}{2}(x - 1)$.

Solution.

- a. To find where the tangent line to the curve is horizontal, we set $\frac{dy}{dx} = 0$, which requires that the numerator be zero, or in other words that

$$(x-1)(x-2) + x(x-2) + x(x-1) = 0.$$

Expanding and combining like terms, we find that $3x^2 - 6x + 2 = 0$, which occurs where $x = \frac{3 \pm \sqrt{5}}{3} \approx 0.42265, 1.57735$. From the graph in Figure 2.7.9, we observe that at each such x -value, there are several corresponding y -values for which the tangent line will be horizontal. For instance, when $x = 0.42265$, then y must satisfy the equation

$$y(y^2 - 1)(y - 2) = 0.42265(0.42265 - 1)(0.42265 - 2) = 0.384900.$$

Because this is a quartic equation (degree 4) equation in y , we use computational technology to help us find the solutions. Doing so, we find four approximate values for y , $y \approx -1.05782, 0.229478, 0.770522, 2.05782$, and thus our estimates for four points at which the tangent line is horizontal are

$$(0.42265, -1.05782); (0.42265, 0.229478); (0.42265, 0.770522); (0.42265, 2.05782).$$

Similar work can be done to find the four points at which the tangent line is horizontal when $x \approx 1.57735$.

- b. The tangent line to the curve is vertical wherever $\frac{dy}{dx}$ is undefined, which occurs precisely where

$$(y^2 - 1)(y - 2) + 2y^2(y - 2) + y(y^2 - 1) = 0.$$

Expanding and combining like terms, we see that we need to solve the cubic equation $4y^3 - 6y^2 - 2y + 2 = 0$; using a computer algebra system, we find that this occurs when $y = \frac{1}{2}, \frac{1 \pm \sqrt{5}}{2}$. It now remains to find the x -coordinate that corresponds to each such y -value. For instance, when $y = \frac{1}{2}$, x must satisfy

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

or in other words, $x^3 - 3x^2 + 2x = \frac{9}{16}$. Here, too, we use technology to determine that there is only one such x , and $x \approx 2.21028$. Similar work can be done to find the x -values that correspond to $y = \frac{1 \pm \sqrt{5}}{2}$.

- c. There are four points on the curve where $x = 1$, which correspond to the y -values that satisfy $y(y^2 - 1)(y - 2) = 0$: $(1, 0), (1, 1), (1, -1), (1, 2)$. We choose the point $(1, 1)$ and evaluate $\frac{dy}{dx}$ at this point. Doing so,

$$\left. \frac{dy}{dx} \right|_{(1,1)} = \frac{(1-1)(1-2) + 1(1-2) + 1(1-1)}{(1^2-1)(1-2) + 2 \cdot 1^2(1-2) + 1(1^2-1)} = \frac{-1}{-2} = \frac{1}{2}.$$

Thus, the equation of the tangent line to the curve at $(1, 1)$ is $y - 1 = \frac{1}{2}(x - 1)$.

Example 2.7.10. For each of the following curves, use implicit differentiation to find dy/dx and determine the equation of the tangent line at the given point.

- a. $x^3 - y^3 = 6xy$, $(-3, 3)$
 b. $\sin(y) + y = x^3 + x$, $(0, 0)$
 c. $3xe^{-xy} = y^2$, $(0.619061, 1)$

Hint.

- a. Note that $\frac{d}{dx}[6xy]$ requires the product rule.
 b. With y being a function of x , $\frac{d}{dx}[\sin(y)]$ requires the chain rule.
 c. To calculate $\frac{d}{dx}[xe^{-xy}]$, first use the product rule and temporarily defer computing $\frac{d}{dx}[e^{-xy}]$.

Answer.

- a. $\frac{dy}{dx}(-3y^2 - 6x) = 6y - 3x^2$ and the tangent line has equation $y - 3 = 1(x + 3)$.
 b. $\frac{dy}{dx} = \frac{3x^2 + 1}{\cos(y) + 1}$. and the tangent line has equation $y = \frac{1}{2}x$.
 c. $\frac{dy}{dx} = \frac{-xye^{-xy} + e^{-xy}}{x^2e^{-xy} + 2}$. and the tangent line is $y - 1 = 0.110794(x - 0.571433)$.

Solution.

- a. Differentiating with respect to x ,

$$\frac{d}{dx}[x^3 - y^3] = \frac{d}{dx}[6xy],$$

so that by the chain and product rules we have

$$3x^2 - 3y^2 \frac{dy}{dx} = 6x \frac{dy}{dx} + 6y.$$

Rearranging to get all terms with $\frac{dy}{dx}$ on the same side, it follows that

$$-3y^2 \frac{dy}{dx} - 6x \frac{dy}{dx} = 6y - 3x^2,$$

and thus

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

Finally, we have established that

$$\frac{dy}{dx} = \frac{6y - 3x^2}{-3y^2 - 6x},$$

so evaluating at $(-3, 3)$, we have $\frac{dy}{dx}|_{(-3,3)} = \frac{6(3) - 3(-3)^2}{-3(3)^2 - 6(-3)} = 1$. Thus, the tangent line has equation $y - 3 = 1(x + 3)$.

b. After differentiating with respect to x , we have

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

Taking the usual steps to solve for $\frac{dy}{dx}$, we find that

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

Evaluating the slope of the tangent line at $(0, 0)$, we have $\frac{dy}{dx} \Big|_{(0,0)} = \frac{1}{2}$, and thus the tangent line at $(0, 0)$ has equation $y = \frac{1}{2}x$.

c. When we differentiate both sides with respect to x ,

$$\frac{d}{dx}[3xe^{-xy}] = \frac{d}{dx}[y^2],$$

we first observe that the product rule is needed on the left and the chain rule on the right. Applying those rules, we have

$$3x \frac{d}{dx}[e^{-xy}] + 3e^{-xy} = 2y \frac{dy}{dx}.$$

Next, we apply the chain rule to differentiate e^{-xy} , which yields

$$3xe^{-xy} \frac{d}{dx}[-xy] + 3e^{-xy} = 2y \frac{dy}{dx}.$$

Finally, to complete the process of differentiation, we use the product rule and get

$$3xe^{-xy}(-x \frac{dy}{dx} - y) + 3e^{-xy} = 2y \frac{dy}{dx}.$$

To solve for $\frac{dy}{dx}$, we first expand to have

$$3x^2e^{-xy} \frac{dy}{dx} - 3xye^{-xy} + 3e^{-xy} = 2y \frac{dy}{dx},$$

and then the usual algebraic work may be done to deduce that

$$\frac{dy}{dx} = \frac{3xye^{-xy} - 3e^{-xy}}{3x^2e^{-xy} - 2y}.$$

Evaluating at the point $(0.571433, 1)$, it follows that the slope of the tangent line is

$$\frac{dy}{dx} \Big|_{(0.571433,1)} = \frac{3 \cdot 0.571433e^{-0.571433} - 3e^{-0.571433}}{3(0.571433)^2e^{-0.571433} - 2} \approx 0.501835.$$

Thus, the tangent line is given by $y - 1 = 0.501835(x - 0.571433)$.

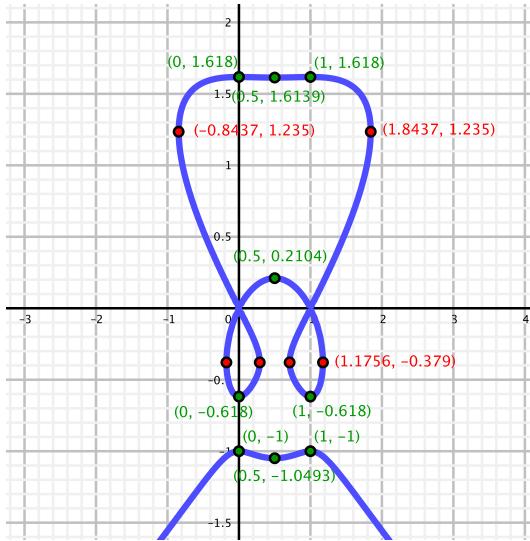
2.7.2 Summary

- In an equation involving x and y where portions of the graph can be defined 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.
- We use implicit differentiation to differentiate an implicitly defined function. We differentiate both sides of the equation with respect to x , treating y as a function of x by applying the chain rule. If possible, we 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 gives the slope of the tangent line to the curve. It may be used to decide where 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.

2.7.3 Exercises

1. (Implicit differentiation in a polynomial equation) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Implicit differentiation in an equation with logarithms) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Implicit differentiation in an equation with inverse trigonometric functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Slope of the tangent line to an implicit curve) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Equation of the tangent line to an implicit curve) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. 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. Be sure to use a graphing utility to plot this implicit curve and to visually check the results of algebraic reasoning that you use to determine where the tangent lines are horizontal and vertical.

Answer. Horizontal tangent lines: $(0, -1), (0, -0.618), (0, 1.618), (1, -1), (1, -0.618), (1, 1.618), (0.5, -1.0493), (0.5, 0.2104), (0.5, 1.6139)$. Vertical tangent lines: $(-0.1756, -0.379), (0.2912, -0.379), (0.7088, -0.379), (1.1756, -0.379), (-0.8437, 1.235)$, and $(1.8437, 1.235)$.



Solution. First, we compute $\frac{dy}{dx}$ by implicit differentiation. Differentiating both sides of the given equation with respect to x while treating y as a function of x , we see that

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

Removing a factor of $\frac{dy}{dx}$ from the lefthand side and dividing to solve for $\frac{dy}{dx}$, we have

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

The tangent line is potentially horizontal where the numerator of $\frac{dy}{dx}$ is zero and potentially vertical where $\frac{dy}{dx}$ is zero. Solving $2x(2x^2 - 3x + 1) = 0$, we see that

$$2x(2x - 1)(x - 1) = 0$$

so the numerator is zero at $x = 0, 0.5, 1$. The denominator is zero where $y(-5y^3 + 6y + 2) = 0$. Using technology to estimate where $-5y^3 + 6y + 2 = 0$, we see the denominator is zero at $y = 0$ and $y \approx -0.856, -0.379, 1.235$.

Next, we need to find the coordinates of these points where potential horizontal or vertical tangent lines occur. First, we consider the potential horizontal tangent line locations. When $x = 0$, $2y^3 + y^2 - y^5 = x^4 - 2x^3 + x^2 = 0$, and thus we know $y^2(2y + 1 - y^3) = 0$, from which it follows that $y = 0$, $y = -1$ or $y \approx -0.618, 1.618$. This leads to four points to consider: $(0, 0)$, $(0, -1)$, $(0, -0.618)$, and $(0, 1.618)$. It is apparent from the graph shown below that the tangent line at $(0, 0)$ is not horizontal (here both the numerator and denominator of $\frac{dy}{dx}$ are zero), but at each of the other three points, the tangent line is horizontal.

We now reason similarly for the other x -values where the numerator is zero. At $x = 1$, $2y^3 + y^2 - y^5 = 1^4 - 2 \cdot 1^3 + 1^2 = 0$, and here again $y = 0$, $y = -1$ or $y \approx -0.618, 1.618$. This leads to four potential points where the tangent line is horizontal: $(1, 0)$, $(1, -1)$, $(1, -0.618)$, and $(1, 1.618)$. At $(1, 0)$, there's not a horizontal tangent line, but at each of the others there is, as seen in the figure. Similar reasoning when $x = 0.5$ shows that $2y^3 + y^2 - y^5 = 0.5^4 - 2 \cdot 0.5^3 + 0.5^2 = 0$.

$0.5^3 + 0.5^2 = 0.625$, from which we find that $y \approx -1.0493, 0.2104, 1.6139$, and thus the three points $(0.5, -1.0493), (0.5, 0.2104), (0.5, 1.6139)$.

Returning to where the denominator is zero ($y = 0$ and $y \approx -0.856, -0.379, 1.235$), we now determine points where the tangent line is vertical. It is apparent from the figure below that there are no vertical tangent lines when $y = 0$ or $y \approx -0.856$, so there are two values to pursue: $y \approx -0.379, 1.235$.

When $y \approx -0.379$, it follows that $2(-0.379)^3 + (-0.379)^2 - (-0.379)^5 = x^4 - 2x^3 + x^2$, so

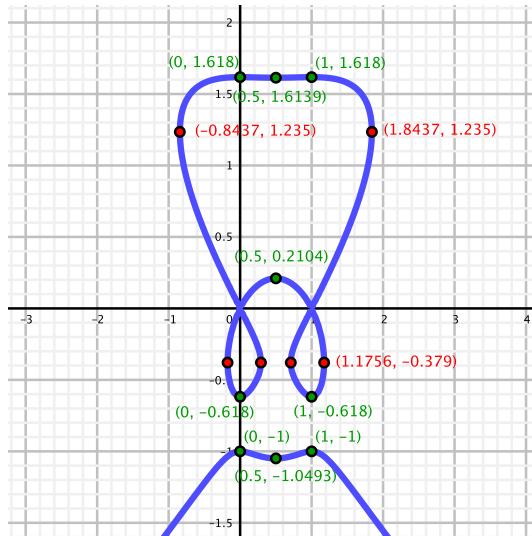
$$x^4 - 2x^3 + x^2 = 0.0426$$

from which we find that $x \approx -0.1756, 0.2912, 0.7088, 1.1756$. The figure confirms that there are vertical tangent lines at $(-0.1756, -0.379), (0.2912, -0.379), (0.7088, -0.379)$, and $(1.1756, -0.379)$. Lastly, when $y \approx 1.235$, it follows $2(1.235)^3 + (1.235)^2 - (1.235)^5 = x^4 - 2x^3 + x^2$, so

$$x^4 - 2x^3 + x^2 = 2.4195$$

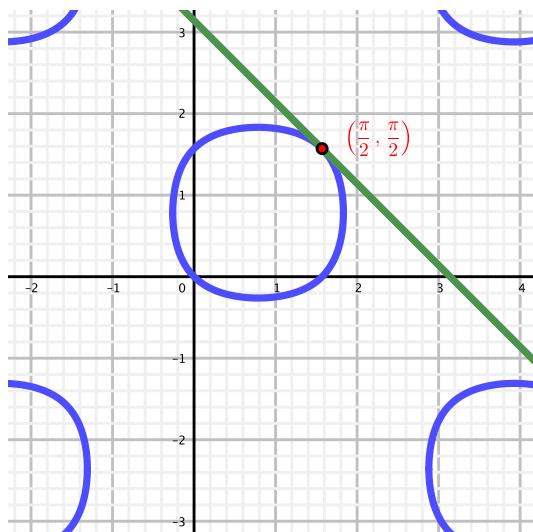
and thus via technology, $x \approx -0.8437, 1.8437$. This leads to two additional points at which the tangent line is vertical: $(-0.8437, 1.235)$ and $(1.8437, 1.235)$.

All of our above work is summarized in the following figure.



7. 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 $(\frac{\pi}{2}, \frac{\pi}{2})$.

Answer. $y = \frac{\pi}{2} - (x - \frac{\pi}{2})$.



Solution. First we need to find the slope of the line tangent to the curve, or $\frac{dy}{dx}$. To do so, we differentiate both sides of the equation with respect to x , treating y as a function of x to obtain

$$\begin{aligned} \frac{d}{dx}(\sin(x+y) + \cos(x-y)) &= \frac{d}{dx}(1) \\ \cos(x+y)\left(1 + \frac{dy}{dx}\right) - \sin(x-y)\left(1 - \frac{dy}{dx}\right) &= 0. \end{aligned}$$

Distributing on the left side of the equation,

$$\cos(x+y) + \cos(x+y)\frac{dy}{dx} - \sin(x-y) + \sin(x-y)\frac{dy}{dx} = 0.$$

Removing a factor of $\frac{dy}{dx}$ from two terms on the right and moving the other two terms to the righthand side,

$$\frac{dy}{dx}(\cos(x+y) + \sin(x-y)) = \sin(x-y) - \cos(x+y).$$

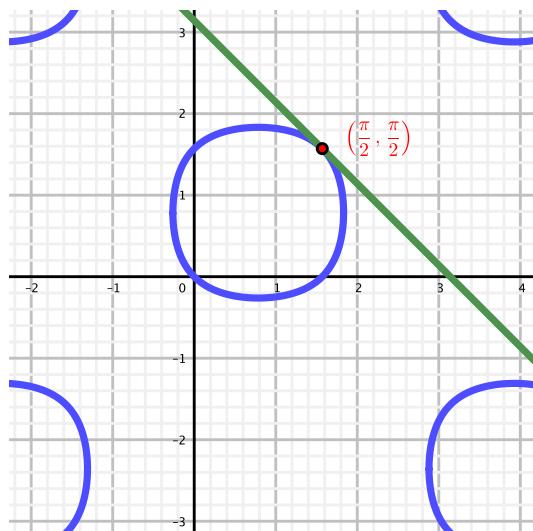
Therefore,

$$\frac{dy}{dx} = \frac{\sin(x-y) - \cos(x+y)}{\cos(x+y) + \sin(x-y)}.$$

At the point $(\frac{\pi}{2}, \frac{\pi}{2})$ we have that

$$\frac{dy}{dx} \Big|_{(\frac{\pi}{2}, \frac{\pi}{2})} = \frac{\sin(0) - \cos(\pi)}{\cos(\pi) + \sin(0)} = -1.$$

Therefore, the equation of the line tangent to the curve at the point $(\frac{\pi}{2}, \frac{\pi}{2})$ is $y = \frac{\pi}{2} - (x - \frac{\pi}{2})$. A portion of the graph of the curve defined by $\sin(x+y) + \cos(x-y) = 1$ and the tangent line to the curve at the point $(\frac{\pi}{2}, \frac{\pi}{2})$ are shown in the following figure.



8. Implicit differentiation enables us a different perspective from which to see why the rule $\frac{d}{dx}[a^x] = a^x \ln(a)$ holds, if we assume that $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$. This exercise leads you through the key steps to do so.

- Let $y = a^x$. Rewrite this equation using the natural logarithm function to write x in terms of y (and the constant a).
- Differentiate both sides of the equation you found in (a) with respect to x , keeping in mind that y is implicitly a function of x .
- Solve the equation you found in (b) for $\frac{dy}{dx}$, and then use the definition of y to write $\frac{dy}{dx}$ solely in terms of x . What have you found?

Answer.

- $x = \frac{\ln(y)}{\ln(a)}$.
- $1 = \frac{1}{\ln(a)} \cdot \frac{1}{y} \frac{dy}{dx}$.
- $\frac{d}{dx}[a^x] = a^x \ln(a)$.

Solution.

- Taking the natural log of both sides, $\ln(y) = \ln(a^x) = x \ln(a)$. Thus, it follows that

$$x = \frac{\ln(y)}{\ln(a)}.$$

- Remembering that $\ln(a)$ is constant with respect to x , we can also write $x = \frac{1}{\ln(a)} \cdot \ln(y)$. Differentiating both sides with respect to x while treating y as a function of x (and again remembering that $\ln(a)$ is constant), we have

$$1 = \frac{1}{\ln(a)} \cdot \frac{1}{y} \frac{dy}{dx}.$$

c. Solving for $\frac{dy}{dx}$, we see that $\frac{dy}{dx} = y \ln(a)$. But recall from the outset that $y = a^x$, and thus

$$\frac{dy}{dx} = a^x \ln(a)$$

which shows that $\frac{d}{dx}[a^x] = a^x \ln(a)$.

2.8 Hyperbolic Functions

Motivating Questions

- What type of function describes the behavior of a line hanging between two poles?
 - What are the properties of a function that can be used to describe the behavior of a line between two poles?
 - What is the derivative of a function that can be used to describe the behavior of a line between two poles?
-

There is an important class of functions that show up in many real-life situations, the so-called hyperbolic functions. Hyperbolic functions can be used to describe the shape of electrical lines freely hanging between two poles or any idealized hanging chain or cable supported only at its ends and hanging under its own weight. Hyperbolic functions can also be used to describe the path of a spacecraft performing a "gravitational slingshot" maneuver.



Figure 2.8.1: Freely-hanging electric power cables can form a catenary.

2.8.1 The Hyperbolic Trigonometric Functions

There are two "fundamental" hyperbolic trigonometric functions, the hyperbolic sine (\sinh) and hyperbolic cosine (\cosh). These functions are defined in terms of the functions e^x and e^{-x} .

Hyperbolic Functions

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \quad \cosh(x) = \frac{e^x + e^{-x}}{2}$$

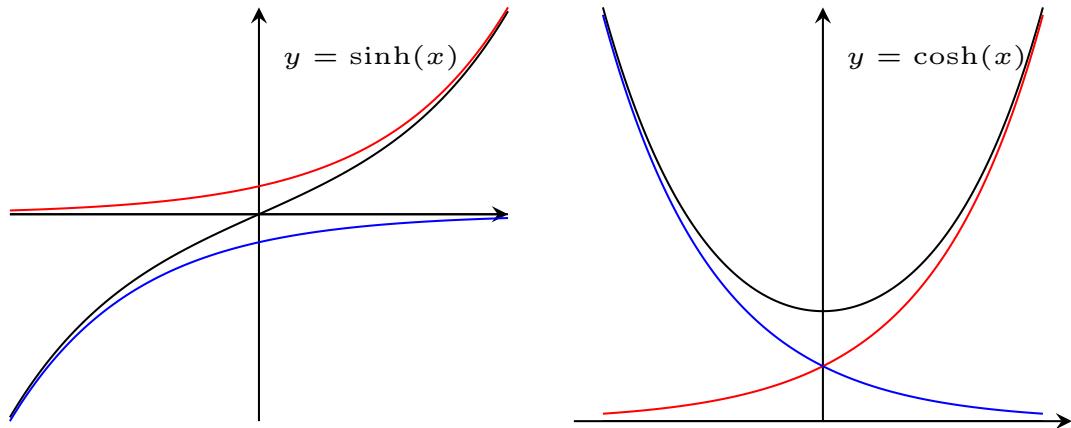


Figure 2.8.2: Graphs of e^x , e^{-x} together with $\sinh(x)$ and $\cosh(x)$.

Example 2.8.3. A cable hanging between two supports will form the shape of a hyperbolic cosine. In particular, the formula

$$y = \frac{T}{w} \cosh\left(\frac{wx}{T}\right),$$

where T is the tension at its lowest point and w is the weight of the cable per unit length, will yield the total cable sag when evaluated at $x = 0$. We can calculate the total sag in a powerline hanging between two poles spaced 400 feet apart where the mass per unit length is 50lb/ft and the tension at the lowest point is 2025lbs . Specifically, the total sag is given by

$$y = \frac{2025\text{lb}}{50\text{lb/ft}} \cosh\left(\frac{50 \times 0}{2025}\right) = 40.5\text{ft}.$$

In addition to the hyperbolic sine and cosine there is also a hyperbolic tangent function which is defined as you might expect.

Hyperbolic Tangent

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

2.8.2 Identities and Properties

Similar to the usual trigonometric functions, the hyperbolic trigonometric functions have several important properties. While we will not take the time to directly show these properties are valid, we do encourage the reader to confirm these properties by inspecting the graphs in Figure 2.8.2 above.

Properties of Hyperbolic Functions

- $\cosh(0) = 1$
- $\sinh(0) = 0$
- $\cosh(-x) = \cosh(x)$
- $\sinh(-x) = -\sinh(x)$

It is also useful to discuss the long-run behavior of the hyperbolic trigonometric functions. Again, inspection of Figure 2.8.2 above suggests that as $x \rightarrow \infty$, the graph of $\cosh(x)$ resembles the graph of $\frac{1}{2}e^x$. Similarly, it appears that as $x \rightarrow -\infty$, the graph of $\cosh(x)$ resembles the graph of $\frac{1}{2}e^{-x}$. This behavior is further explained by using the formulas for $\cosh(x)$ and $\sinh(x)$ and the facts that $e^{-x} \rightarrow 0$ as $x \rightarrow \infty$, and $e^x \rightarrow 0$ as $x \rightarrow -\infty$.

Recall that the trig functions were defined on the unit circle and hence were called the *circular functions*. This is where we get the Pythagorean identity, setting $x = \cos \theta$ and $y = \sin \theta$,

$$\cos^2 \theta + \sin^2 \theta = 1 \text{ because } x^2 + y^2 = 1.$$

In fact, an analogous identity holds for the hyperbolic trigonometric functions.

$$\cosh^2 \theta - \sinh^2 \theta = 1$$

This identity shows us how the hyperbolic functions got their name. Suppose (x, y) is a point in the plane, and $x = \cosh \theta$ and $y = \sinh \theta$ for some θ . Then the point (x, y) lies on the hyperbola $x^2 - y^2 = 1$.

2.8.3 Derivatives of Hyperbolic Functions

We now proceed to calculate the derivatives of each of the hyperbolic functions.

Example 2.8.4. Calculate the derivatives

1. $\frac{d}{dx} \cosh(x)$
2. $\frac{d}{dx} \sinh(x)$
3. $\frac{d}{dx} \tanh(x)$

Hint. Recall that $\frac{d}{dx} e^x = e^x$, $\sinh(x) = \frac{e^x - e^{-x}}{2}$, $\cosh(x) = \frac{e^x + e^{-x}}{2}$, and $\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

Solution.

$$\frac{d}{dx} \cosh(x) = \frac{d}{dx} \left(\frac{e^x + e^{-x}}{2} \right) = \left(\frac{e^x - e^{-x}}{2} \right) = \sinh(x)$$

$$\frac{d}{dx} \sinh(x) = \frac{d}{dx} \left(\frac{e^x - e^{-x}}{2} \right) = \left(\frac{e^x + e^{-x}}{2} \right) = \cosh(x)$$

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

Derivatives of Hyperbolic Trigonometric Functions

$$1. \frac{d}{dx} \cosh(x) = \sinh(x)$$

$$2. \frac{d}{dx} \sinh(x) = \cosh(x)$$

$$3. \frac{d}{dx} \tanh(x) = \frac{1}{\cosh^2(x)}$$

2.8.4 Applications of Hyperbolic Trigonometry

A company wishes to build a suspension bridge that stretches between the basketball arena and the baseball stadium on the other side of the railway lines in a particular city. The center part of the bridge will be suspended between two concrete pillars 280 feet apart and 80 feet high. The cable holding the bridge is to be exactly 30 feet above the railway tracks in the middle of the bridge, i.e., it sags exactly 50 feet.

Gottfried Leibniz and Christian Huygens in 1691 determined that any cable hanging under the force of gravity must have the shape of the graph of

$$y(x) = a \cosh \left(\frac{x}{a} \right) + b.$$

This shape is known as a *catenary*. The parameter a is the ratio of cable tension to cable density. The only use of the parameter b is to provide a vertical shift, if needed.

We can ask two important questions. First, what values must a and b have in order for the catenary to fit the constraints provided by the placement of the concrete pillars and the low point of the cable?

In order to find a and b we need to solve two separate equations. We know that $y(0) = 30$ to ensure we have sufficient clearance above the railway tracks. We also know that $y(140) = 80$ since the cable attaches to a 80 foot tall pillar 140 feet from the lowest point (center). Therefore we have,

$$30 = a \cosh \left(\frac{0}{a} \right) + b$$

$$80 = a \cosh\left(\frac{140}{a}\right) + b$$

Since $\cosh(0) = 1$ we can simplify the first equation to $30 = a + b$ substituting this equation into the second equation yields $80 = a \cosh\left(\frac{140}{a}\right) + 30 - a$. Now, by graphing the $y = x \cosh\left(\frac{140}{x}\right) + 30 - x$ and the line $y = 80$ and then graphically searching for the point of intersection (as in Figure 2.8.5) we see that $80 = a \cosh\left(\frac{140}{a}\right) + 30 - a$ when $a \approx 203.82$.

Geogebra: www.geogebra.org/m/pcj3yebb

Figure 2.8.5: Solving $80 = a \cosh\left(\frac{140}{a}\right) + 30 - a$

Using the value of $a \approx 203.82$ together with $30 = a + b$ we have $b = -173.82$. Therefore, the height of the bridge can be modeled by the equation

$$h(x) = 203.82 \cosh\left(\frac{x}{203.82}\right) - 173.82$$

2.8.5 Exercises

1. (Simplifying Hyperbolic Trigonometric Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Limits of Hyperbolic Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Proportionality) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Limits of Hyperbolic Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Graphs of Hyperbolic Functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

2.9 The Tangent Line Approximation

Motivating 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?
-

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 $(a, f(a))$. If we know both a point on the line and the slope of the line we can find the equation of the tangent line and write the equation in point-slope form¹.

Example 2.9.1. Consider the function $y = g(x) = -x^2 + 3x + 2$.

- Use the limit definition of the derivative to compute a formula for $y = g'(x)$.
- Determine the slope of the tangent line to $y = g(x)$ at the value $x = 2$.
- Compute $g(2)$.
- Find an equation for the tangent line to $y = g(x)$ at the point $(2, g(2))$. Write your result in point-slope form.
- On the axes provided in Figure 2.9.2, sketch an accurate, labeled graph of $y = g(x)$ along with its tangent line at the point $(2, g(2))$.

¹Recall 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.

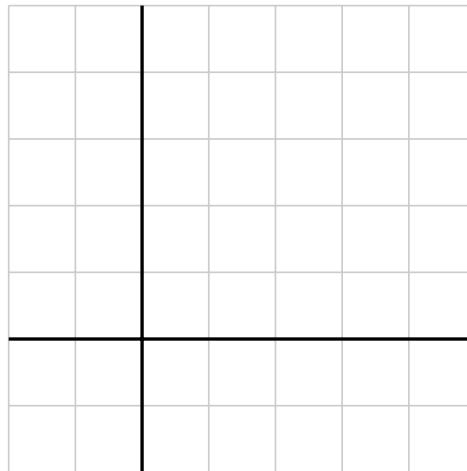


Figure 2.9.2: Axes for plotting $y = g(x)$ and its tangent line to the point $(2, g(2))$.

2.9.1 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 equation of the resulting tangent line is given in point-slope form by

$$y - f(a) = f'(a)(x - a) \text{ or } y = f'(a)(x - a) + f(a).$$

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 2.9.3, we see 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. The function and its tangent line are nearly indistinguishable up close. Local linearity can also be seen dynamically in the java applet at <http://gvsu.edu/s/6J>.

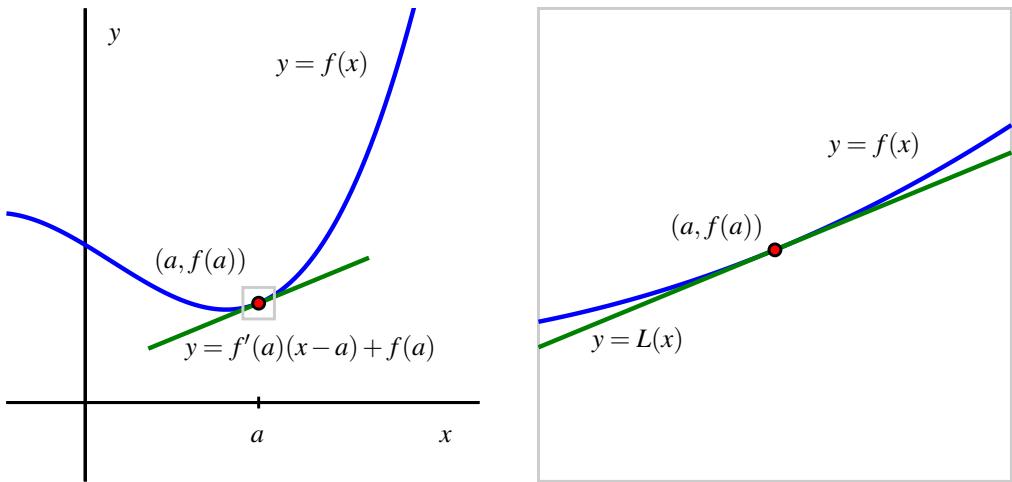


Figure 2.9.3: A function $y = f(x)$ and its tangent line at the point $(a, f(a))$: at left, from a distance, and at right, up close. At right, we label the tangent line function by $y = L(x)$ and observe that for x near a , $f(x) \approx L(x)$.

2.9.2 The local linearization

A slight change in perspective and notation will enable us to be more precise in discussing how the tangent line approximates f near $x = a$. By solving for y , we can write the equation for the tangent line as

$$y = f'(a)(x - a) + f(a)$$

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, $L(x)$ is nothing more than a new name for the tangent line. As we saw above, for x close to a , $f(x) \approx L(x)$.

Example 2.9.4. Suppose 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 . To estimate 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

$$f(1.2) \approx L(1.2) = 3 - 2(1.2 - 1) = 3 - 2(0.2) = 2.6.$$

We emphasize that $y = L(x)$ is simply a new name for the tangent line function. Using this new notation and our observation that $L(x) \approx f(x)$ for x near a , it follows that we can write

$$f(x) \approx f(a) + f'(a)(x - a) \text{ for } x \text{ near } a.$$

Example 2.9.5. 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 $L(x) = -2 + 3(x + 1)$.

- Compute the values of $L(-1)$ and $L'(-1)$.
- What must be the values of $g(-1)$ and $g'(-1)$? Why?
- Do you expect the value of $g(-1.03)$ to be greater than or less than the value of $g(-1)$? Why?
- Use the local linearization to estimate the value of $g(-1.03)$.
- Suppose that you also know that $g''(-1) = 2$. What does this tell you about the graph of $y = g(x)$ at $a = -1$?
- 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 2.9.6.

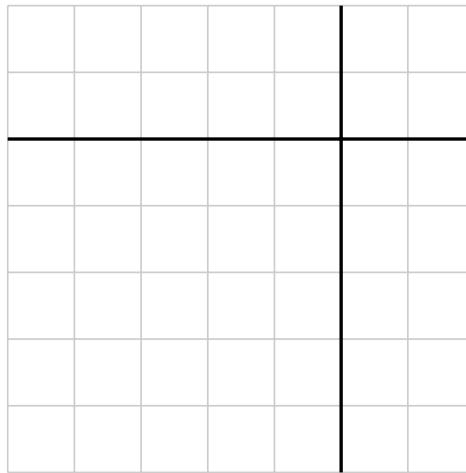


Figure 2.9.6: Axes for plotting $y = L(x)$ and $y = g(x)$.

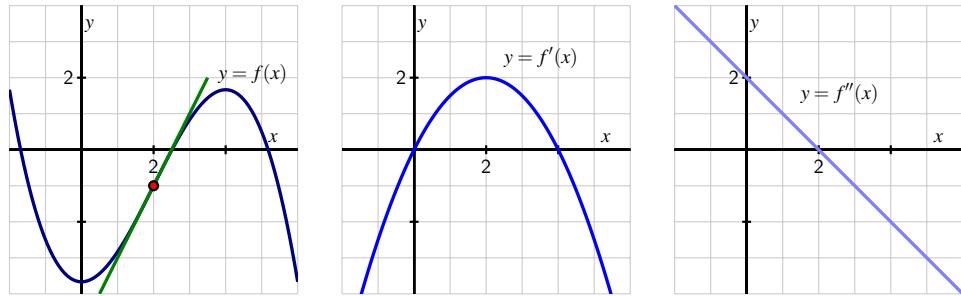
Hint.

- Follow the rule for L .
- Recall that the form of the local linearization is $L(x) = g(a) + g'(a)(x - a)$.
- Is the function g increasing or decreasing at $a = -1$?
- Remember that $g(-1.03) \approx L(-1.03)$.
- What does the second derivative tell you about the shape of a curve?
- Use your work above.

Answer.

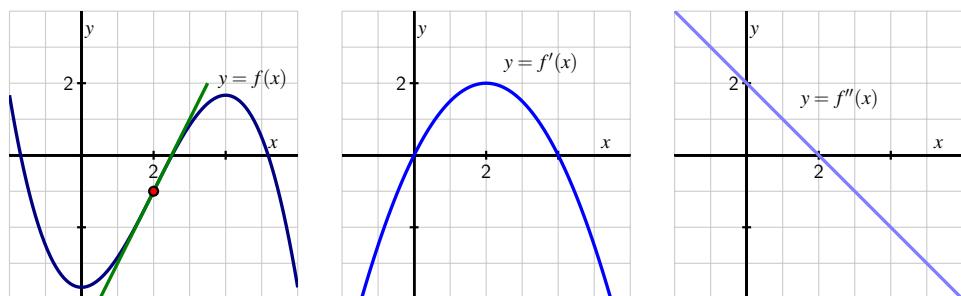
- $L(-1) = -2$; $L'(-1) = 3$.

- b. $g(-1) = -2$; $g'(-1) = 3$.
- c. Increasing.
- d. $Og(-1.03) \approx L(-1.03) = -2.09$.
- e. Concave up.
- f. The illustration below shows a possible graph of $y = g(x)$ near $x = -1$, along with the tangent line $y = L(x)$ through $(-1, g(-1))$.



Solution.

- a. Using the formula for L , $L(-1) = -2$; Since $L'(x) = 3$, we see $L'(-1) = 3$.
- b. Since $L(x) = g(-1) + g'(-1)(x + 1) = -2 + 3(x + 1)$, we see $g(-1) = -2$ and $g'(-1) = 3$. Alternatively, we could observe that the value and slope of g must match the value and slope of L at the point of tangency.
- c. Because $g'(-1) = 3$, we see that g is increasing at $a = -1$?
- d. Observe that $g(-1.03) \approx L(-1.03) = -2 + 3(-1.03 + 1) = -2 - 0.09 = -2.09$.
- e. Since $g''(-1) > 0$, we know g is concave up at $x = -1$.
- f. In the figure below, we use the results of our previous work to generate the plot shown, which is a possible graph of $y = g(x)$ near $x = -1$, along with the tangent line $y = L(x)$ through $(-1, g(-1))$.



From Example 2.9.5, we see that 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,

- because $L(x) = f(a) + f'(a)(x - a)$, it follows that $L(a) = f(a)$; and
- because 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 $(a, f(a))$.

Thus, if we know the linear approximation $y = L(x)$ for a function, we know the original function's value and its 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 shown in Figure 2.9.7.

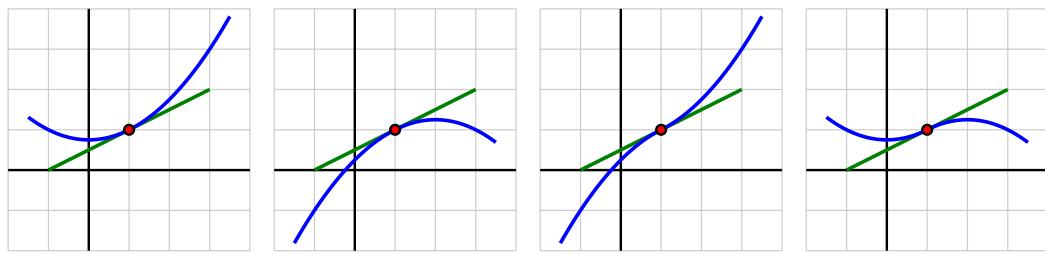


Figure 2.9.7: Four possible graphs for a nonlinear differentiable function and how it can be situated relative to its tangent line at a point.

These possible shapes result 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 f is concave down and the tangent line lies above the curve, as shown in the second figure.
- If $f''(a) = 0$ and f'' changes sign at $x = a$, the concavity of the graph will change, and we will see either the third or fourth figure.²
- A fifth option (which is not very interesting) can occur if the function f itself is linear, so that $f(x) = L(x)$ for all values of x .

The plots in Figure 2.9.7 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

²It is possible that $f''(a) = 0$ but f'' does not change sign at $x = a$, in which case the graph will look like one of the first two options.

instance, in the first situation in the leftmost plot in Figure 2.9.7 where $f''(a) > 0$, because the tangent line falls below the curve, we know that $L(x) \leq f(x)$ for all values of x near a .

Example 2.9.8. This example 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 2.9.9

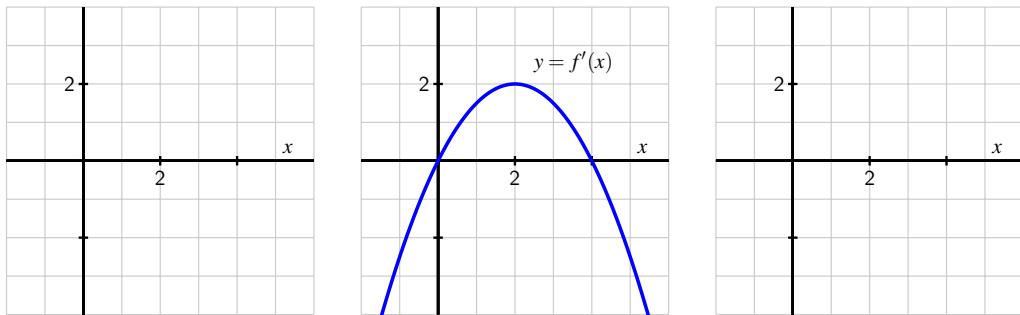


Figure 2.9.9: At center, a graph of $y = f'(x)$; at left, axes for plotting $y = f(x)$; at right, axes for plotting $y = f''(x)$.

Your task is to determine as much information as possible about f (especially near the value $a = 2$) by responding to the questions below.

- Find a formula for the tangent line approximation, $L(x)$, to f at the point $(2, -1)$.
- Use the tangent line approximation to estimate the value of $f(2.07)$. Show your work carefully and clearly.
- Sketch a graph of $y = f''(x)$ on the righthand grid in Figure 2.9.9; label it appropriately.
- Is the slope of the tangent line to $y = f(x)$ increasing, decreasing, or neither when $x = 2$? Explain.
- Sketch a possible graph of $y = f(x)$ near $x = 2$ on the lefthand grid in Figure 2.9.9. 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.
- Does your estimate in (b) over- or under-estimate the true value of $f(2.07)$? Why?

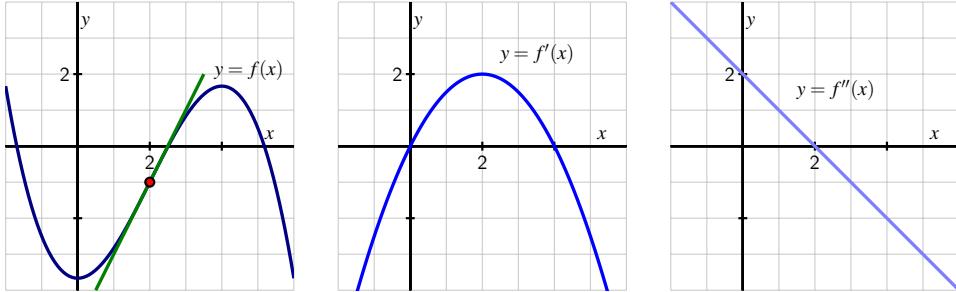
Hint.

- Find the value of $f'(2)$ from the given graph of f .

- b. Remember that $f(2.07) \approx L(2.07)$.
- c. The graph of $y = f''(x)$ is the derivative of the graph of $y = f'(x)$.
- d. Is f' increasing, decreasing, or neither when $x = 2$?
- e. Draw $y = L(x)$ first. Then think about options for f relative to the graph of L .
- f. Does the tangent line lie above or below the graph of $y = f(x)$ at $(2, 3)$?

Answer.

- a. $L(x) = -1 + 2(x - 2)$.
- b. $f(2.07) \approx L(2.07) = -0.86$.
- c. See the image in part e.
- d. Neither.
- e. See the image below, which shows, at left, a possible graph of $y = f(x)$ near $x = 2$, along with the tangent line $y = L(x)$ through $(2, f(2))$.

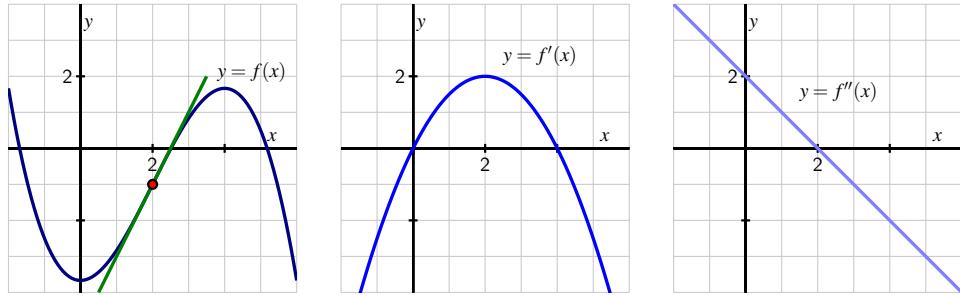


- f. Too large.

Solution.

- a. Since $f(2) = -1$ and $f'(2) = 2$, we have $L(x) = -1 + 2(x - 2)$.
- b. Using our work in (a), $f(2.07) \approx L(2.07) = -1 + 2(2.07 - 2) = -1 + 2 \cdot 0.07 = -0.86$.
- c. See the image in part e.
- d. The slope of the tangent line to $y = f(x)$ is increasing for $x < 2$ because $y = f'(x)$ is an increasing function on this interval. Similarly, for $x > 2$, the slope of the tangent line to $y = f(x)$ is decreasing. Right at $x = 2$, the slope of the tangent line to $y = f(x)$ is neither increasing nor decreasing.

- e. See the plot in the image below, which shows, at left, a possible graph of $y = f(x)$ near $x = 2$, along with the tangent line $y = L(x)$ through $(2, f(2))$. Note that $y = f(x)$ is concave up for $x < 2$ since f' is increasing on that interval, and $y = f(x)$ is concave down for $x > 2$ since f' is decreasing there. Hence $y = f(x)$ changes from concave up to concave down right at $x = 2$, which is also the point near 2 where the graph of $y = f(x)$ is steepest.



- f. Because the tangent line to $y = f(x)$ lies above the graph of f to the right of $x = 2$, our estimate of $f(2.07)$ is too large — the local linearization overshoots the true value of f at this point.

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 certain challenging limits. For instance, we have seen that the limit

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x}$$

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, $\sin(x) \approx x$, and therefore

$$\frac{\sin(x)}{x} \approx \frac{x}{x} = 1,$$

which makes plausible the fact that

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1.$$

2.9.3 Summary

- 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

$$y - f(a) = f'(a)(x - a).$$

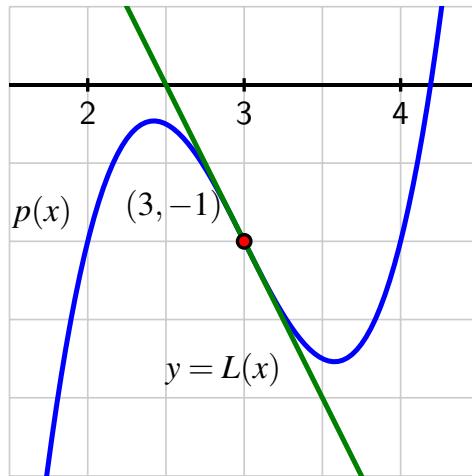
- The principle of local linearity tells us that if we zoom in on a point where a function $y = f(x)$ is differentiable, the function will be indistinguishable from its tangent line. That is, a differentiable function looks linear when viewed up close. We rename the tangent line to be the function $y = L(x)$, where $L(x) = f(a) + f'(a)(x - a)$. Thus, $f(x) \approx L(x)$ for all x near $x = a$.
- If we know the tangent line approximation $L(x) = f(a) + f'(a)(x - a)$ to a function $y = f(x)$, then because $L(a) = f(a)$ and $L'(a) = f'(a)$, we also know the values of both the function and its derivative 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 .

2.9.4 Exercises

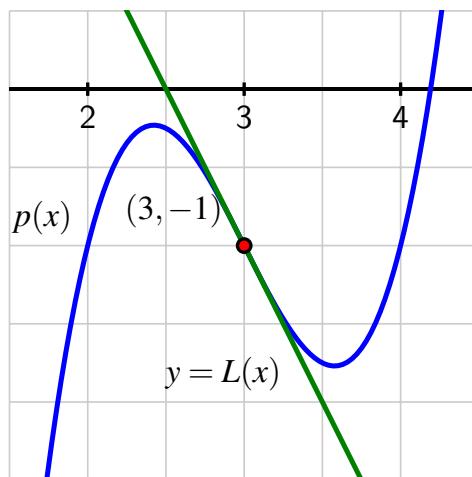
1. (Approximating \sqrt{x}) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Local linearization of a graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Estimating with the local linearization) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Predicting behavior from the local linearization) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. 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.

Answer.

- a. $p(3) = -1$ and $p'(3) = -2$.
- b. $p(2.79) \approx -0.58$.
- c. Too large.
- d.

**Solution.**

- Since the value and slope of the tangent line match the value and slope of the function at the point of tangency, we know $p(3) = L(3) = -1$ and $p'(3) = L'(3) = -2$.
- Using the local linearization, $p(2.79) \approx L(2.79) = -2(2.79) + 5 = -0.58$.
- Since $p''(x) < 0$ for $x < 3$, we know that p is concave down for $x < 3$. This means to the left of the point of tangency at $x = 3$, the tangent line lies above the curve, and thus the estimate in (b) is too large.
- In the following figure, we see a possible function p that matches the value and slope of the tangent line at the point $(3, -1)$ and that is concave down to the left of $x = 3$ and concave up to the right.



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

Table 2.9.10: Temperature data for the potato.

- Use a central difference to estimate $F'(60)$. Use this estimate as needed in subsequent questions.
- Find the local linearization $y = L(t)$ to the function $y = F(t)$ at the point where $a = 60$.
- Determine an estimate for $F(63)$ by employing the local linearization.
- Do you think your estimate in (c) is too large or too small? Why?

Answer.

- $F'(60) \approx 1.56$ degrees per minute.
- $L(t) \approx 1.56(t - 60) + 324.5$.
- $F(63) \approx L(63) \approx 329.18$ degrees F.
- Overestimate.

Solution.

- The central difference approximation with $h = 15$ gives

$$F'(60) \approx \frac{F(75) - F(45)}{30} = \frac{342.8 - 296}{30} = 1.56$$

degrees per minute.

- The local linearization at $a = 60$ is $L(t) = F'(60)(t - 60) + F(60) \approx 1.56(t - 60) + 324.5$.
- Using the linearization from (b), $F(63) \approx L(63) \approx 1.56(3) + 324.5 = 329.18$ degrees F.
- Since the values of $F(t)$ given in the table appear to be increasing at a decreasing rate, it follows that the graph of F is concave down. This implies that near $t = 60$, the graph of $L(t)$ (the tangent line to F at $t = 60$) lies above the graph of F . Therefore, the approximation given in part (c) is probably an overestimate.
- An object moving along a straight line path has a differentiable position function $y = s(t)$; $s(t)$ measures the object's position relative to the origin at time t . It is known that at time $t = 9$ seconds, the object's position is $s(9) = 4$ feet (i.e., 4 feet to the right of the origin). 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.
 - Use local linearity to estimate the position of the object at $t = 9.34$.
 - 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 the origin or away from it? Is its velocity increasing or decreasing?

Answer.

- $s(9.34) \approx L(9.34) = 3.592$.
- underestimate.
- The object is slowing down as it moves toward toward its starting position at $t = 4$.

Solution.

- The linearization L of s at $t = a$ is $L(t) = s(a) + s'(a)(t - a)$. With the information we have we can find the linearization of s at $t = 9$, using the facts that $s(9) = 4$ and $s'(9) = -1.2$. Thus, the linearization of s at $t = 9$ is $L(t) = 4 - 1.2(t - 9)$. One use of the linearization is that for t close to 9, $L(t)$ can provide a good approximation to $s(t)$. So

$$s(9.34) \approx L(9.34) = 4 - 1.2(0.34) = 3.592.$$

- Since $s''(4) = 0.08$ is positive, we know that the graph of s is concave up at $t = 4$. This means that the linearization of s at $t = 4$ lies below the graph of s and so $L(9.34)$ is an underestimate to $s(9.34)$.
- Given that $s'(4)$ is negative, the object is moving back toward its original position. The fact that $s''(4)$ is positive means that the velocity of the object is increasing at $t = 4$. This means that the object is actually slowing down as it moves toward toward its starting position at $t = 4$.
- 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 $y = f(x)$.
 - At what x -value(s) is $f'(x) = 0$? Justify your answer algebraically, but include a graph of f' to support your conclusion.
 - 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.
 - 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.

Answer.

- $x = 1$.
- On $-0.37 < x < 1.37$; f is concave up.
- $f(1.88) \approx 0.02479$, and this estimate is larger than the true value of $f(1.88)$.

Solution.

- Since e raised to any power is always positive (and thus never 0), it follows that the only value that makes $f'(x) = (x - 1)e^{-x^2} = 0$ is $x = 1$. This is confirmed by plotting f' where we only see the graph cross the x -axis at $x = 1$ (though we do see the graph approach the x -axis) as x increases or decreases without bound).
- We know that $f''(x) > 0$ wherever f' is increasing. From the graph of f' , we see that f' is increasing approximately on the interval $-0.37 < x < 1.37$, and thus f is concave up on this interval.

- c. The tangent line approximation at $a = 2$ is $L(x) = f(2) + f'(2)(x - 2) \approx 0.0182 + -0.05495(x - 2)$. Thus, $f(1.88) \approx L(1.88) \approx 0.0182 + -0.05495(1.88 - 2) = 0.02479$. Since $f''(x) < 0$ at and around $x = 2$, f is concave down there, and therefore the tangent line lies above the curve. This makes our estimate larger than the true value of $f(1.88)$.

2.10 The Mean Value Theorem

Motivating Questions

- Can we use global information about a function (average rate of change) to say anything about the local behavior of the function (instantaneous rate of change)?
-

If you know how a function behaves at every point then you know quite a lot about a function. In fact, this is the whole idea behind plotting points on a coordinate plane in order to determine what the graph of a function looks like. Indeed, given enough local information about a function we can describe it very well globally. However, the opposite is not obviously true. If you know how a function behaves overall, can you say anything about how the function behaves at a specific point? For example, if you knew that $f(x)$ was a parabola, can you determine the value of $f(3)$? No! In order to determine the value of $f(3)$ you need more information. In general, global information is not enough to determine local information. However, there are some special cases and one such special case is the mean value theorem.

2.10.1 The Mean Value Theorem

Example 2.10.1. Imagine you have driven a car between Lincoln and Omaha Nebraska, a distance of 59 miles. Suppose that it took you exactly three quarters of an hour to drive this distance. Is it possible that you drove less than 75 miles per hour the entire trip? The average rate of change was 78.67 miles per hour our intuition tells us that at some point on the trip you must have driven at least 78.67 miles per hour. In fact, this is exactly the mean value theorem.

The Mean Value Theorem

Let f be a continuous function on the interval $[a, b]$ and differentiable on the interval (a, b) , then there exists a number c , with $a < c < b$, such that

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

In other words, $f(b) - f(a) = f'(c)(b - a)$.

A proof of the mean value theorem is outside the scope of the text book. However, it is worth noting that the mean value theorem essentially states that given any secant line between two points on a continuous curve, there is a line tangent to the curve at a point in-between those two points with the property that the slope of the tangent line and slope of the secant line are identical. A geometric representaiton of the theorem is located in Figure 2.10.2.

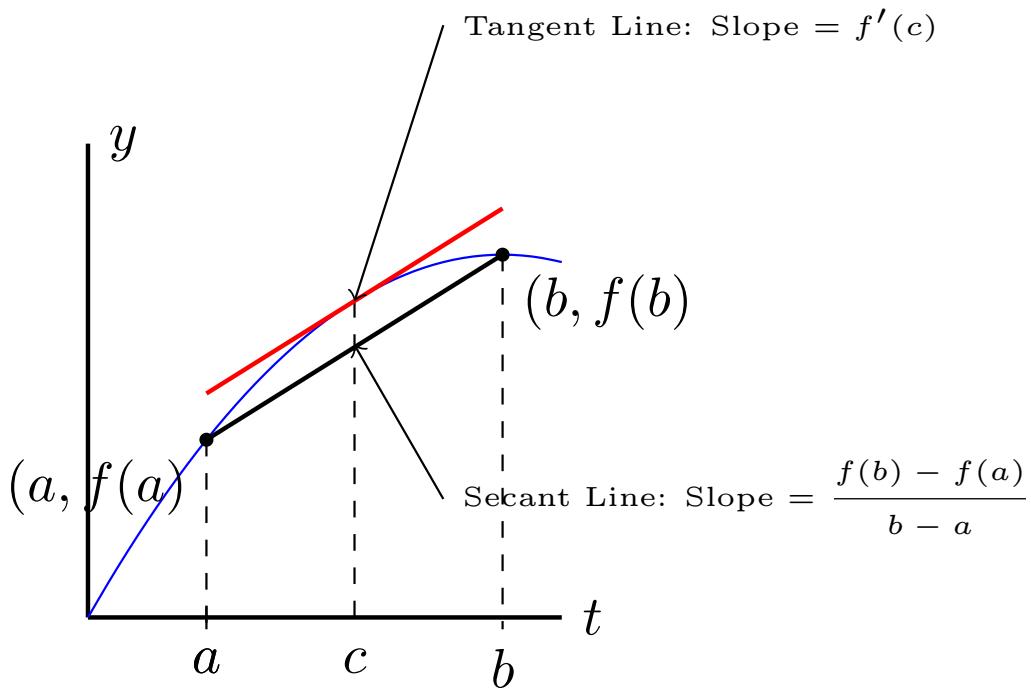


Figure 2.10.2: To understand this theorem geometrically, look at the figure. Join the points on the curve where $x = a$ and $x = b$ with a secant line and observe that the slope of secant line $(f(b) - f(a))/(b - a)$. Notice that there appears to be at least one point between a and b where the slope of the tangent line to the curve is precisely the same as the slope of the secant line.

Example 2.10.3.

2.10.2 Consequences of the Mean Value Theorem

The Mean Value Theorem has several important consequences that confirm some intuitive ideas that you probably already hold. First, the *Increasing Function Theorem* establishes that continuous functions with a positive derivative on an interval are increasing on that interval. More specifically,

The Increasing Function Theorem

Suppose that f is continuous on $a \leq x \leq b$ and differentiable on $a < x < b$.

- If $f'(x) > 0$ on $a < x < b$, then f is increasing on $a \leq x \leq b$.
- If $f'(x) \geq 0$ on $a < x < b$, then f is non-decreasing on $a \leq x \leq b$.

We can also state that a function with a derivative equal to zero on an interval is constant on that interval.

Constant Function Theorem

Suppose that f is continuous on $a \leq x \leq b$ and differentiable on $a < x < b$. If $f'(x) = 0$ on $a < x < b$, then f is constant on $a \leq x \leq b$.

Example 2.10.4.

2.10.3 Exercises

1. (Statement of the Mean Value Theorem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Applying Theorems) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Conclusion of the Mean Value Theorem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Conclusion of the Mean Value Theorem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

Chapter 3

Using Derivatives

3.1 Using derivatives to identify extreme values

Motivating Questions

- What are the critical numbers 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?
-

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 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 3.1.2) that represents the height of an object tossed vertically: its maximum value occurs at the vertex of the parabola and represents the greatest height the object reaches. This maximum value is an especially important point on the graph, the point at which the curve changes from increasing to decreasing.

Definition 3.1.1. Given a function f , we say that $f(c)$ is a **global** or **absolute maximum** of f provided that $f(c) \geq f(x)$ for all x in the domain of f , and similarly we call $f(c)$ a **global** or **absolute minimum** of f whenever $f(c) \leq f(x)$ for all x in the domain of f .

For instance, in Figure 3.1.2, 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. 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 minima or maxima.

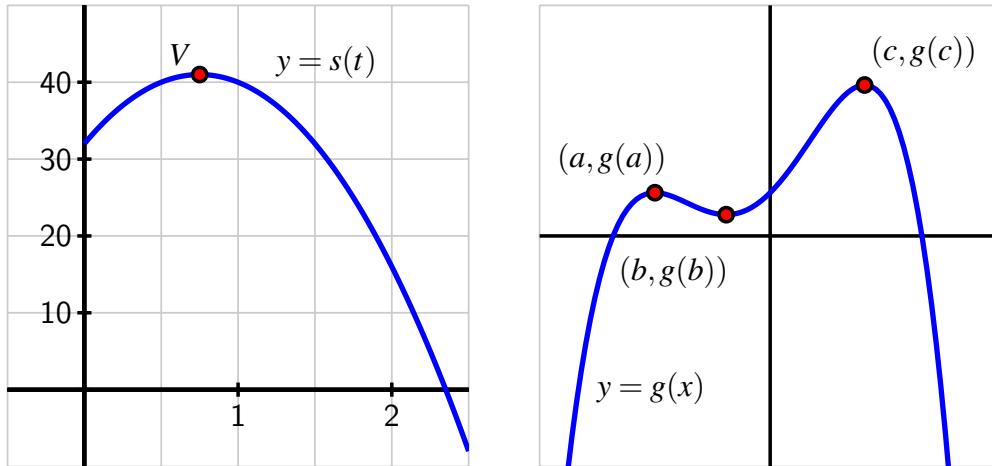


Figure 3.1.2: At left, $s(t) = -16t^2 + 24t + 32$ whose vertex is $(\frac{3}{4}, 41)$; at right, a function g that demonstrates several high and low points.

Definition 3.1.3. We say that $f(c)$ is a **local maximum** or **relative maximum** of f provided that $f(c) \geq f(x)$ for all x near c , and $f(c)$ is called a **local** or **relative minimum** of f whenever $f(c) \leq f(x)$ for all x near c .

For example, in Figure 3.1.2, 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 $g(c)$; it can also be considered a relative maximum. Any maximum or minimum may also be called an *extreme value* of f .

We would like to use calculus ideas to 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.

Example 3.1.4. Consider the function h given by the graph in Figure 3.1.5. Use the graph to answer each of the following questions.

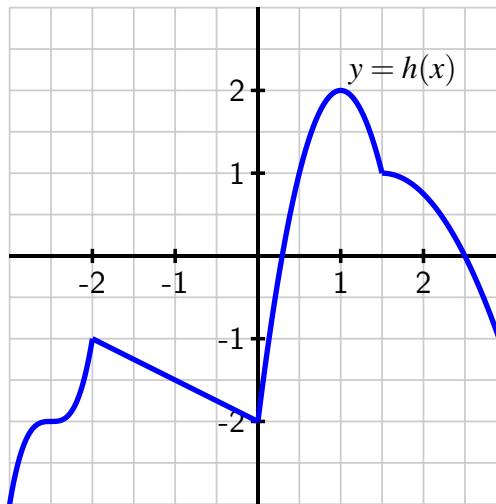


Figure 3.1.5: The graph of a function h on the interval $[-3, 3]$.

- a. Identify all of the values of c for which $h(c)$ is a local maximum of h .
- b. Identify all of the values of c for which $h(c)$ is a local minimum of h .
- c. Does h have a global maximum on the interval $[-3, 3]$? If so, what is the value of this global maximum?
- d. Does h have a global minimum on the interval $[-3, 3]$? 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.

3.1.1 Critical numbers and the first derivative test

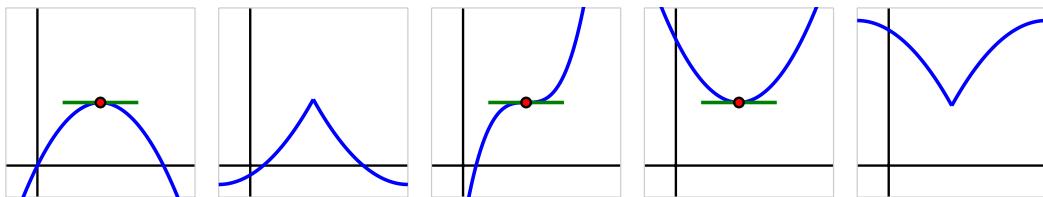


Figure 3.1.6: From left to right, a function with a relative maximum where its derivative is zero; a function with a relative maximum where its derivative is undefined; a function with neither a maximum nor a minimum at a point where its derivative is zero; a function with a relative minimum where its derivative is zero; and a function with a relative minimum where its derivative is undefined.

If a continuous function has a relative maximum at c , then it is both necessary and sufficient that the function change from being increasing just before c to decreasing just after c . A continuous function has a relative minimum at c if and only if the function changes from decreasing to increasing at c . (See Figure 3.1.6.) 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 numbers*.

Definition 3.1.7. We say that a function f has a **critical number** at $x = c$ provided that c is in the domain of f , and $f'(c) = 0$ or $f'(c)$ is undefined.

Critical numbers are the only possible locations where the function f may have relative extremes. Note that not every critical number produces a maximum or minimum; in the middle graph of Figure 3.1.6, 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 number does not yield a maximum or minimum.

When c is a critical number, we say that $(c, f(c))$ is a *critical point* of the function, or that $f(c)$ is a *critical value*. The *first derivative test* summarizes how sign changes in the first derivative (which can only occur at critical numbers) indicate the presence of a local maximum or minimum for a given function.

First Derivative Test

If p is a critical number 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 .

Example 3.1.8. Let f be a function whose derivative is given by the formula $f'(x) = e^{-2x}(3-x)(x+1)^2$. Determine all critical numbers of f and decide whether a relative maximum, relative minimum, or neither occurs at each.

Solution. Since we already have $f'(x)$ written in factored form, it is straightforward to find

the critical numbers of f . Because $f'(x)$ 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 numbers of f . (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 inputs near the critical numbers. Because the critical numbers 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 numbers: for instance, the sign of f' must be the same for every $x < -1$. We create a first derivative sign chart to summarize the sign of f' on the relevant intervals, along with the corresponding behavior of f .

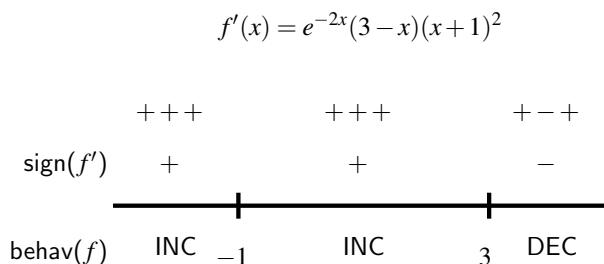


Figure 3.1.9: The first derivative sign chart for a function f whose derivative is given by the formula $f'(x) = e^{-2x}(3-x)(x+1)^2$.

To produce the first derivative sign chart in Figure 3.1.9 we identify the sign of each factor of $f'(x)$ at one selected point in each interval. For instance, for $x < -1$, we could 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 results in a positive value for f' , which we denote by the “+” in the interval to the left of $x = -1$. And, since f' is positive on that interval, we know that f is increasing, so we write “INC” to represent the 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 we look for critical numbers at which f' changes sign. In this example, f' changes sign only at $x = 3$, from positive to negative, so f has a relative maximum at $x = 3$. Although f has a critical number 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 3.1.10. Suppose that $g(x)$ is a function continuous for every value of $x \neq 2$ whose first derivative is $g'(x) = \frac{(x+4)(x-1)^2}{x-2}$. Further, assume that it is known that g has a vertical asymptote at $x = 2$.

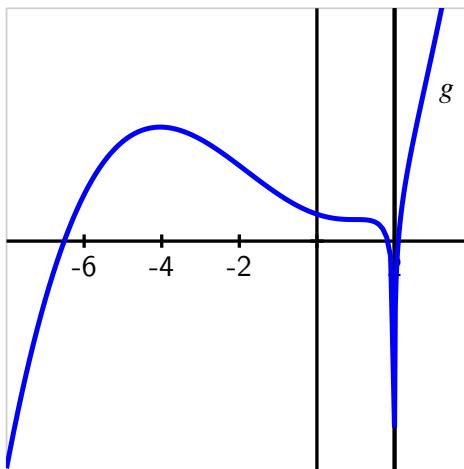
- Determine all critical numbers of g .
- By developing a carefully labeled first derivative sign chart, decide whether g has as a local maximum, local minimum, or neither at each critical number.
- Does g have a global maximum? global minimum? Justify your claims.
- What is the value of $\lim_{x \rightarrow \infty} g'(x)$? What does the value of this limit tell you about the long-term behavior of g ?
- Sketch a possible graph of $y = g(x)$.

Hint.

- For which x is $g'(x) = 0$?
- Note that $(x - 1)^2$ is positive for all $x \neq 1$.
- Use your first derivative sign chart from (b).
- Try expanding the numerator before evaluating the limit.
- Think carefully about the information from the sign chart found in (b).

Answer.

- $x = -4$ or $x = 1$.
- g has a local maximum at $x = -4$ and neither a max nor min at $x = 1$.
- g does not have a global minimum; it is unclear (at this point in our work) if g increases without bound, so we can't say for certain whether or not g has a global maximum.
- $\lim_{x \rightarrow \infty} g'(x) = \infty$.
- A possible graph of g is the following.



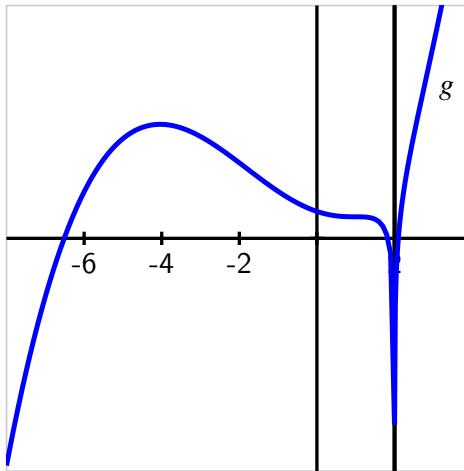
Solution.

- a. Since $g'(x) = \frac{(x+4)(x-1)^2}{x-2}$, we see that $g'(x) = 0$ implies that $x = -4$ or $x = 1$. While $x = 2$ makes g' undefined, we are told that g has a vertical asymptote at $x = 2$, so $x = 2$ is not in the domain of g , and hence is technically not a critical number of g . Nonetheless, we place $x = 2$ on our first derivative sign chart since the vertical asymptote is a location at which g' may change sign.
- b. The first derivative sign chart shows that $g'(x) > 0$ for $x < -4$, $g'(x) < 0$ for $-4 < x < 1$, $g'(x) < 0$ for $1 < x < 2$, and $g'(x) > 0$ for $x > 2$. By the first derivative test, g has a local maximum at $x = -4$ and neither a max nor min at $x = 1$. As these are the only two critical numbers, these are the only two locations for possible extremes. (Note: although g changes from decreasing to increasing at $x = 2$, this is due to a vertical asymptote, and g does not have a minimum there.)
- c. Because g is decreasing as $x \rightarrow 2^-$ (where g has a vertical asymptote), g does not have a global minimum. For $x > 2$, g is always increasing, which suggests that g does not have a global maximum (though we do not know for sure that g increases without bound).
- d. We observe that

$$\begin{aligned}
 \lim_{x \rightarrow \infty} g'(x) &= \lim_{x \rightarrow \infty} \frac{(x+4)(x-1)^2}{x-2} \\
 &= \lim_{x \rightarrow \infty} \frac{x^3 + 2x^2 - 7x + 4}{x-2} \cdot \frac{\frac{1}{x}}{\frac{1}{x}} \\
 &= \lim_{x \rightarrow \infty} \frac{x^2 + 2x - 7 + \frac{4}{x}}{1 - \frac{2}{x}} \\
 &= \infty
 \end{aligned}$$

Since $g'(x) \rightarrow \infty$ as $x \rightarrow \infty$, this tells us that g increases without bound as $x \rightarrow \infty$.

- e. From all of our work above, we know that g has a local maximum at $x = -4$, a horizontal tangent line with neither a max nor min at $x = 1$, and a vertical asymptote at $x = 2$, plus g and g' both increase without bound as $x \rightarrow \infty$. Thus, a possible graph of g is the following.



3.1.2 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, so throughout that interval the tangent line to $y = f(x)$ lies below the curve at every point. At a point where $f'(p) = 0$, the sign of the second derivative determines whether f has a local minimum or local maximum at the critical number p .

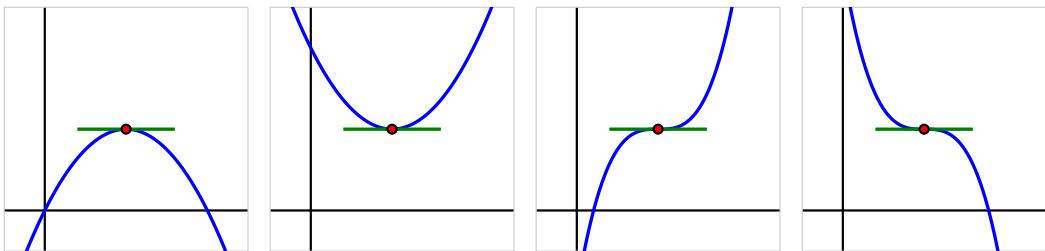


Figure 3.1.11: Four possible graphs of a function f with a horizontal tangent line at a critical point.

In Figure 3.1.11, we see the four possibilities for a function f that has a critical number 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 number, 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 f is concave down at p with a horizontal tangent line, so f has a local maximum there. This fact, along with the corresponding statement for when $f''(p)$ is positive, is the substance of the *second derivative test*.

Second Derivative Test

If p is a critical number 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.¹

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 3.1.12. 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 numbers 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 numbers of f . The first derivative sign chart for f is given in Figure 3.1.13.

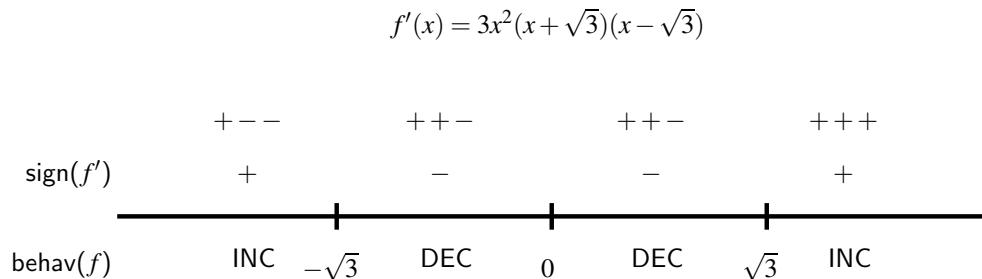


Figure 3.1.13: The first derivative sign chart for f when $f'(x) = 3x^4 - 9x^2 = 3x^2(x^2 - 3)$.

We see that f is increasing on the intervals $(-\infty, -\sqrt{3})$ and $(\sqrt{3}, \infty)$, and f is decreasing on $(-\sqrt{3}, 0)$ and $(0, \sqrt{3})$. By the first derivative test, this information tells us that f has a local maximum at $x = -\sqrt{3}$ and a local minimum at $x = \sqrt{3}$. Although f also has a critical number 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

¹Consider the functions $f(x) = x^4$, $g(x) = -x^4$, and $h(x) = x^3$ at the critical point $p = 0$.

and negative, we first find where $f''(x) = 0$. Observe that

$$0 = 12x^3 - 18x = 12x \left(x^2 - \frac{3}{2} \right) = 12x \left(x + \sqrt{\frac{3}{2}} \right) \left(x - \sqrt{\frac{3}{2}} \right).$$

This equation has solutions $x = 0, \pm\sqrt{\frac{3}{2}}$. Building a sign chart for f'' in the exact same way we do for f' , we see the result shown in Figure 3.1.14.

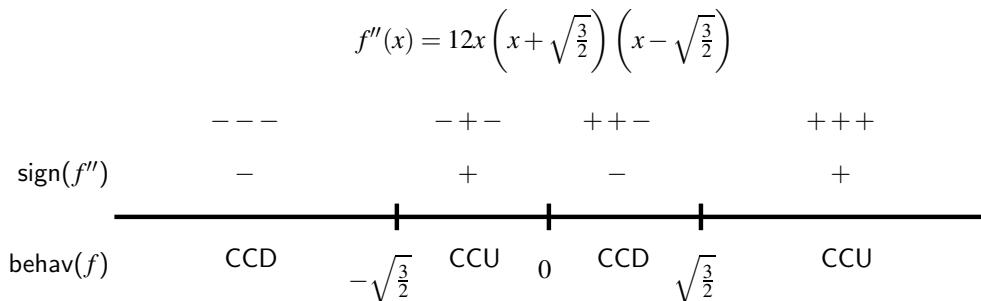


Figure 3.1.14: The second derivative sign chart for f when $f''(x) = 12x^3 - 18x = 12x^2 \left(x^2 - \sqrt{\frac{3}{2}} \right)$.

Therefore, f is concave down on the intervals $(-\infty, -\sqrt{\frac{3}{2}})$ and $(0, \sqrt{\frac{3}{2}})$, and concave up on $(-\sqrt{\frac{3}{2}}, 0)$ and $(\sqrt{\frac{3}{2}}, \infty)$.

Putting all of this information together, we now see a complete and accurate possible graph of f in Figure 3.1.15.

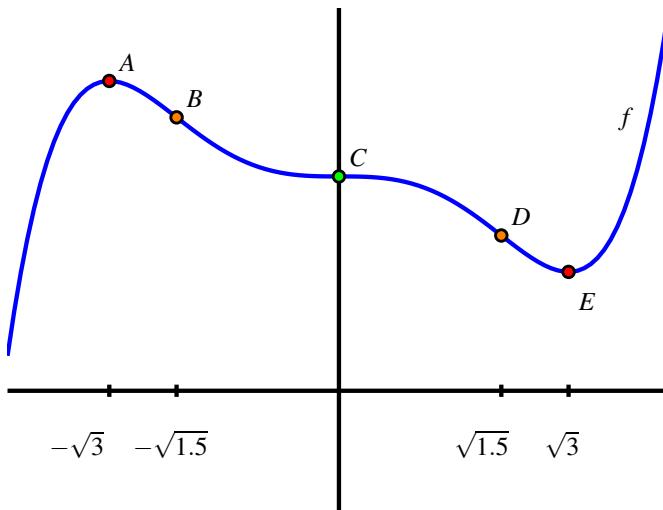


Figure 3.1.15: A possible graph of the function f in Example 3.1.12.

The point $A = (-\sqrt{3}, f(-\sqrt{3}))$ is a local maximum, because 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 3.1.14. Finally, at point C , f has a critical point 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 3.1.15 would satisfy the given criteria for f .

Points B , C , and D in Figure 3.1.15 are locations at which the concavity of f changes. We give a special name to any such point.

Definition 3.1.16. 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** (or **point of inflection**) 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.

At this point in our study, it is important 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.

Example 3.1.17. Suppose that g is a function whose second derivative, g'' , is given by the graph in Figure 3.1.18.

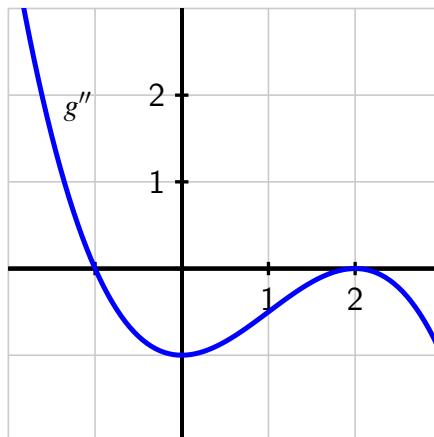


Figure 3.1.18: The graph of $y = g''(x)$.

- Find the x -coordinates of all points of inflection of g .
- Fully describe the concavity of g by making an appropriate sign chart.
- 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 number of g , or is it impossible to say? Why?
- Assuming that $g''(x)$ is a polynomial (and that all important behavior of g'' is seen in the graph above), what degree polynomial do you think $g(x)$ is? Why?

Hint.

- What must be true of $g''(x)$ at a point of inflection?
- Use the given graph to decide where g'' is positive and negative.
- What does the second derivative test say?
- Can you guess a formula for $g''(x)$ based on its graph?

Answer.

- $x = -1$ is an inflection point of g .
- g is concave up for $x < -1$, concave down for $-1 < x < 2$, and concave down for $x > 2$.

- c. g has a local minimum at $x = -1.67857351$.
- d. g is a degree 5 polynomial.

Solution.

- a. Based on the given graph of g'' , the only point at which g'' changes sign is $x = -1$, and hence this is an inflection point of g .
- b. Note that $g''(x) > 0$ for $x < -1$, $g''(x) < 0$ for $-1 < x < 2$, and $g''(x) < 0$ for $x > 2$. This tells us that g is concave up for $x < -1$, concave down for $-1 < x < 2$, and concave down for $x > 2$.
- c. Given that $g'(-1.67857351) = 0$, we know that g has a horizontal tangent line at this critical number. In addition, from the given graph of g'' , we see that $g''(-1.67857351) > 0$ and observe that g is concave up at x -values near -1.67857351 . By the second derivative test, g has a local minimum at $x = -1.67857351$.
- d. From the given graph, since g'' has a simple zero at $x = -1$ and a repeated zero at $x = 2$, it appears that g'' is a degree 3 polynomial. If so, then g' is a degree 4 polynomial, and g is a degree 5 polynomial.

As we will see in more detail in the following section, derivatives also help us to understand families of functions that differ only by changing one or more parameters. For instance, we might be interested in understanding the behavior of all functions of the form $f(x) = a(x - h)^2 + k$ where a , h , and k are parameters. Each parameter has considerable impact on how the graph appears.

Example 3.1.19. Consider the family of functions given by $h(x) = x^2 + \cos(kx)$, where k is an arbitrary positive real number.

- a. Use a graphing utility to sketch the graph of h for several different k -values, including $k = 1, 3, 5, 10$. Plot $h(x) = x^2 + \cos(3x)$ on the axes provided. What is the smallest value of k at which you think you can see (just by looking at the graph) at least one inflection point on the graph of h ?

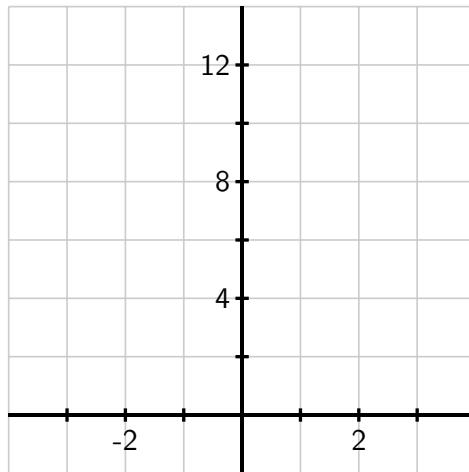


Figure 3.1.20: Axes for plotting $y = h(x)$.

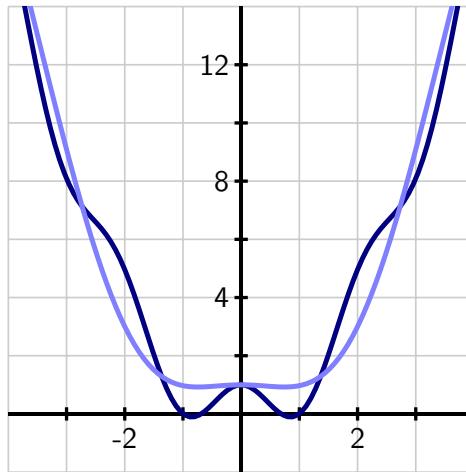
- b. Explain why the graph of h has no inflection points if $k \leq \sqrt{2}$, but infinitely many inflection points if $k > \sqrt{2}$.
- c. Explain why, no matter the value of k , h can only have finitely many critical numbers.

Hint.

- a. Be sure to try some values between 1.5 and 2.
- b. Treat k as an arbitrary constant in computing $h'(x)$ and $h''(x)$.
- c. What is the largest value of $k \sin(kx)$? How many times can $y = k \sin(kx)$ intersect the line $y = 2x$?

Answer.

- a. In the graph below, $h(x) = x^2 + \cos(3x)$ is given in dark blue, while $h(x) = x^2 + \cos(1.6x)$ is shown in light blue.



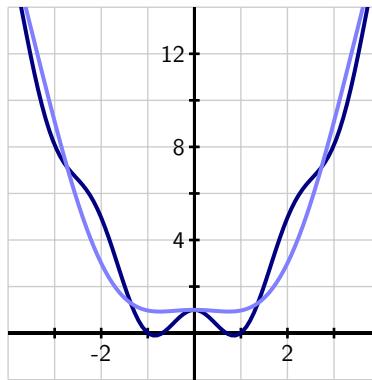
- b. If $\frac{2}{k^2} > 1$, then the equation $\cos(kx) = \frac{2}{k^2}$ has no solution. Hence, whenever $k^2 < 2$, or $k < \sqrt{2} \approx 1.414$, it follows that the equation $\cos(kx) = \frac{2}{k^2}$ has no solutions x , which means that $h''(x)$ is never zero (indeed, for these k -values, $h''(x)$ is always positive so that h is always concave up). On the other hand, if $k \geq \sqrt{2}$, then $\frac{2}{k^2} \leq 1$, which guarantees that $\cos(kx) = \frac{2}{k^2}$ has infinitely many solutions, due to the periodicity of the cosine function. At each such point, $h''(x) = 2 - k^2 \cos(kx)$ changes sign, and therefore h has infinitely many inflection points whenever $k \geq \sqrt{2}$.
- c. To see why h can only have a finite number of critical numbers regardless of the value of k , consider the equation

$$0 = h'(x) = 2x - k \sin(kx),$$

which implies that $2x = k \sin(kx)$. Since $-1 \leq \sin(kx) \leq 1$, we know that $-k \leq k \sin(kx) \leq k$. Once $|x|$ is sufficiently large, we are guaranteed that $|2x| > k$, which means that for large x , $2x$ and $k \sin(kx)$ cannot intersect. Moreover, for relatively small values of x , the functions $2x$ and $k \sin(kx)$ can only intersect finitely many times since $k \sin(kx)$ oscillates a finite number of times. This is why h can only have a finite number of critical numbers, regardless of the value of k .

Solution.

- a. In the graph below, $h(x) = x^2 + \cos(3x)$ is given in dark blue, while $h(x) = x^2 + \cos(1.6x)$ is shown in light blue.



Close inspection of the light blue graph reveals some subtle changes in concavity around $x \approx \pm 0.5$. For values smaller than 1.6, it is very hard to visually detect any inflection points in $h(x)$.

- b. Treating k as an arbitrary constant, we first observe that $h'(x) = 2x - k \sin(kx)$. Again treating k as a constant and differentiating, we find

$$h''(x) = 2 - k^2 \cos(kx).$$

We seek the values of x for which $h''(x) = 0$ at which h'' changes sign. Setting $h''(x) = 0$ and rearranging the resulting equation, we now seek x such that

$$k^2 \cos(kx) = 2,$$

or

$$\cos(kx) = \frac{2}{k^2}.$$

Now, remember that k is an arbitrary positive constant and recall that $-1 \leq \cos(\theta) \leq 1$ for all input values θ . If $\frac{2}{k^2} > 1$, then the equation $\cos(kx) = \frac{2}{k^2}$ has no solution. Hence, whenever $k^2 < 2$, or $k < \sqrt{2} \approx 1.414$, it follows that the equation $\cos(kx) = \frac{2}{k^2}$ has no solutions x , which means that $h''(x)$ is never zero (indeed, for these k -values, $h''(x)$ is always positive so that h is always concave up). On the other hand, if $k \geq \sqrt{2}$, then $\frac{2}{k^2} \leq 1$, which guarantees that $\cos(kx) = \frac{2}{k^2}$ has infinitely many solutions, due to the periodicity of the cosine function. At each such point, $h''(x) = 2 - k^2 \cos(kx)$ changes sign, and therefore h has infinitely many inflection points whenever $k \geq \sqrt{2}$.

- c. To see why h can only have a finite number of critical numbers regardless of the value of k , consider the equation

$$0 = h'(x) = 2x - k \sin(kx),$$

which implies that $2x = k \sin(kx)$. Since $-1 \leq \sin(kx) \leq 1$, we know that $-k \leq k \sin(kx) \leq k$. Once $|x|$ is sufficiently large, we are guaranteed that $|2x| > k$, which means that for large x , $2x$ and $k \sin(kx)$ cannot intersect. Moreover, for relatively small values of x , the functions $2x$ and $k \sin(kx)$ can only intersect finitely many times since $k \sin(kx)$ oscillates a finite number of times. This is why h can only have a finite number of critical numbers, regardless of the value of k .

3.1.3 Summary

- The critical numbers 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, the sign of f'' tells us the concavity of f and hence 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.

3.1.4 Exercises

1. (Finding critical points and inflection points) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Finding inflection points) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Matching graphs of f , f' , f'') This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. This problem concerns a function about which the following information is known:
 - f is a differentiable function defined at every real number x
 - $f(0) = -1/2$
 - $y = f'(x)$ has its graph given at center in Figure 3.1.21

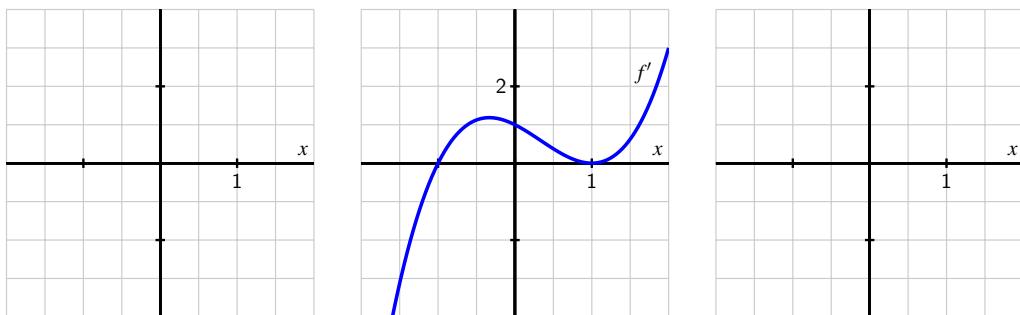
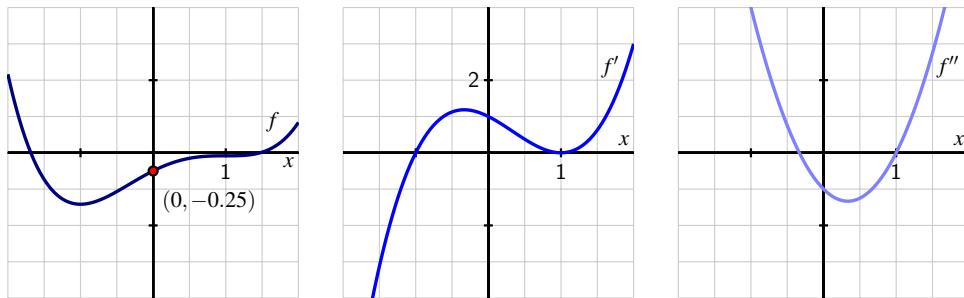


Figure 3.1.21: At center, a graph of $y = f'(x)$; at left, axes for plotting $y = f(x)$; at right, axes for plotting $y = f''(x)$.

- Construct a first derivative sign chart for f . Clearly identify all critical numbers of f , where f is increasing and decreasing, and where f has local extrema.
- On the right-hand axes, sketch an approximate graph of $y = f''(x)$.
- Construct a second derivative sign chart for f . Clearly identify where f is concave up and concave down, as well as all inflection points.
- On the left-hand axes, sketch a possible graph of $y = f(x)$.

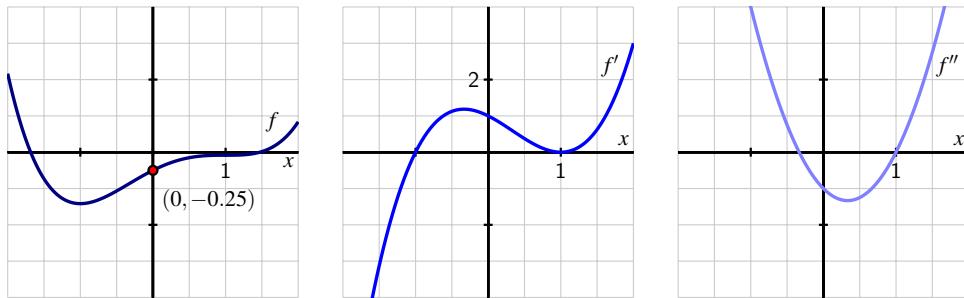
Answer.

- f' is positive for $-1 < x < 1$ and for $x > 1$; f' is negative for all $x < -1$. f has a local minimum at $x = -1$.
- A possible graph of $y = f''(x)$ is shown at right in the figure.
- $f''(x)$ is negative for $-0.35 < x < 1$; $f''(x)$ is positive everywhere else; f has points of inflection at $x \approx -0.35$ and $x = 1$.
- A possible graph of $y = f(x)$ is shown at left in the figure.

**Solution.**

- From the given graph of f' , the first derivative sign chart for f shows that f has critical numbers at $x = -1, 1$ and that f' is positive for $-1 < x < 1$ and for $x > 1$, while being negative for all $x < -1$. Hence f is increasing for $-1 < x < 1$ and for $x > 1$, and decreasing for $x < -1$. It follows by the first derivative test that f has a local minimum at $x = -1$ because f changes from decreasing to increasing there.
- Based on where $f'(x)$ has horizontal tangent lines and where its slope is positive and negative, we can sketch an approximate graph of $y = f''(x)$ in the usual way that we sketch the derivative of a given function. From the given graph of $y = f'(x)$, we can see that $f''(x) = 0$ at $x \approx -0.35$ and at $x = 1$, and that between these points, $f''(x)$ is negative because $f'(x)$ is decreasing, while $f''(x)$ is positive everywhere else since $f'(x)$ is increasing everywhere else. A possible graph of $y = f''(x)$ is shown at right in the following figure.
- From our sketch of the second derivative in (b), we can develop a second derivative sign chart for f . In particular, we saw that $f''(x) = 0$ at $x \approx -0.35$ and at $x = 1$, and that between these points, $f''(x)$ is negative, while $f''(x)$ is positive everywhere else. This shows that f is concave down for $-0.35 < x < 1$ while being concave up everywhere else. We thus see that f has points of inflection at both $x \approx -0.35$ and $x = 1$.

- d. Now that we know where f is increasing and decreasing as well as where f is concave up and concave down, along with the fact that $f(0) = -0.5$, we can sketch an approximate graph of $y = f(x)$. Note that we also know f has a horizontal tangent line at $x = 1$ while being increasing on either side of $x = 1$, and earlier we established that f has a local minimum at $x = -1$. After making a few rough sketches and shifting the graph vertically to pass through $(0, -0.5)$, we arrive at the graph shown at left below.



5. Suppose that g is a differentiable function and $g'(2) = 0$. In addition, suppose that on $1 < x < 2$ and $2 < x < 3$ it is known that $g'(x)$ is positive.
- Does g have a local maximum, local minimum, or neither at $x = 2$? Why?
 - Suppose that $g''(x)$ exists for every x such that $1 < x < 3$. Reasoning graphically, describe the behavior of $g''(x)$ for x -values near 2.
 - Besides being a critical number of g , what is special about the value $x = 2$ in terms of the behavior of the graph of g ?

Answer.

- Neither.
- $g''(2) = 0$; g'' is negative for $1 < x < 2$ and positive for $2 < x < 3$.
- g has a point of inflection at $x = 2$.

Solution.

- g has neither a local maximum nor local minimum at $x = 2$ because f' does not change sign at $x = 2$.
 - Because $g'(2) = 0$ and g' is positive for $1 < x < 2$ and for $2 < x < 3$, it must be that g' is decreasing for $1 < x < 2$ and g' is increasing for $2 < x < 3$. Since g' changes from decreasing to increasing at $x = 2$, it follows that g' has a local minimum at $x = 2$ and thus $g''(2) = 0$. In addition, $g''(x)$ must change sign from negative to positive at $x = 2$. In particular, g'' is negative for $1 < x < 2$ and positive for $2 < x < 3$.
 - g has a horizontal tangent line at $x = 2$, is increasing before $x = 2$ and decreasing after $x = 2$, and concave down before $x = 2$ and concave up after $x = 2$, and thus g has a point of inflection at $x = 2$.
6. Suppose that h is a differentiable function whose first derivative is given by the graph in Figure 3.1.22.

- How many real number solutions can the equation $h(x) = 0$ have? Why?
- If $h(x) = 0$ has two distinct real solutions, what can you say about the signs of the two solutions? Why?
- Assume that $\lim_{x \rightarrow \infty} h'(x) = 3$, as appears to be indicated in Figure 3.1.22. How will the graph of $y = h(x)$ appear as $x \rightarrow \infty$? Why?
- Describe the concavity of $y = h(x)$ as fully as you can from the provided information.

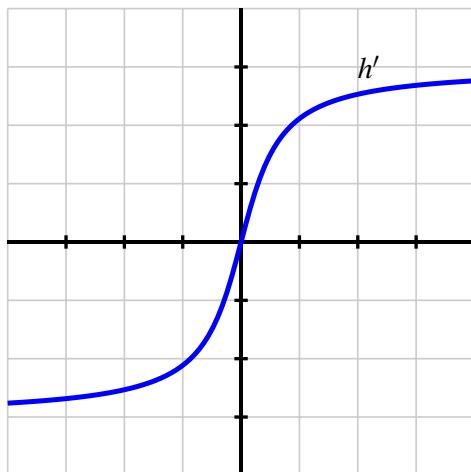
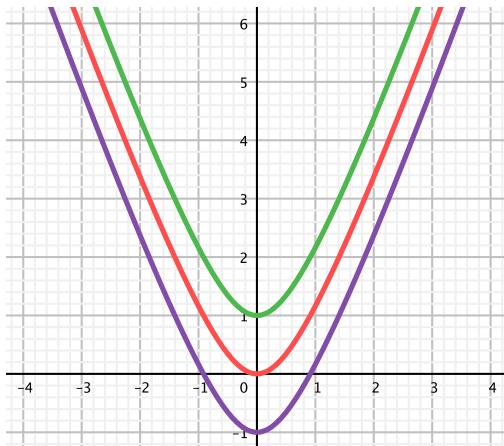


Figure 3.1.22: The graph of $y = h'(x)$.

Answer.

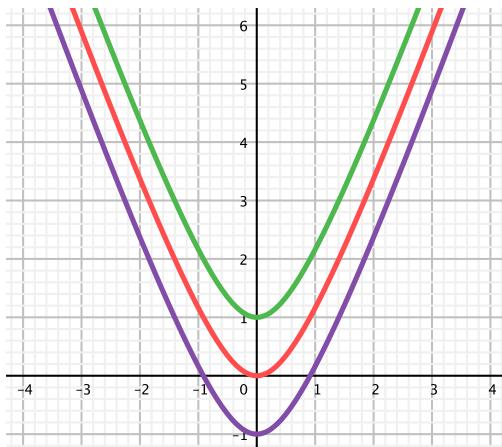
- h can have no, one, or two real zeros.



- One root is negative and the other positive.
- h will look like a line with slope 3.
- h is concave up everywhere; h is almost linear for large values of $|x|$.

Solution.

- Since h' is negative to the left of 0 and positive to the right of 0, the function h is decreasing for $x < 0$ and increasing for $x > 0$. So h can have at most 2 real roots (solutions to $h(x) = 0$). As the graphs in the following figure show, it is possible that h has no real zeros, one real zero, or two real zeros.



- b. Given that h is decreasing for $x < 0$ and increasing for $x > 0$, the function h has an absolute minimum value at $x = 0$. So one root is to the left of this minimum and the other to the right. Thus, one root is negative and the other positive. See, for instance, the lowest of the three possibilities for h shown in the figure in (a).
- c. The slopes of the tangent lines to h will all be near 3, but less than 3, as x gets large. So the graph of h will look very much like a line with slope 3 for large values of x .
- d. We note that h' is increasing everywhere, h is concave up everywhere. Since the graph of h' levels off for large values of $|x|$, the concavity of h is small when $|x|$ is large. Also, the largest value of the slope of the tangent line to h' occurs when $x = 0$, so h has its largest second derivative when $x = 0$. So h is almost linear as x decreases without bound, h has more bend as x approaches 0, has its greatest bend at $x = 0$, then the second derivative of h as x increases, ultimately having h become almost linear as x increases without bound.
7. Let p be a function whose second derivative is $p''(x) = (x + 1)(x - 2)e^{-x}$.
- Construct a second derivative sign chart for p and determine all inflection points of p .
 - Suppose you also know that $x = \frac{\sqrt{5}-1}{2}$ is a critical number of p . Does p have a local minimum, local maximum, or neither at $x = \frac{\sqrt{5}-1}{2}$? Why?
 - If the point $(2, \frac{12}{e^2})$ lies on the graph of $y = p(x)$ and $p'(2) = -\frac{5}{e^2}$, find the equation of the tangent line to $y = p(x)$ at the point where $x = 2$. Does the tangent line lie above the curve, below the curve, or neither at this value? Why?

Answer.

- $p''(x)$ is negative for $-1 < x < 2$ and positive for all other values of x ; p has points of inflection at $x = -1$ and $x = 2$.
- Local maximum.
- Neither.

Solution.

- Note that $p''(x) = 0$ at $x = -1$ and $x = 2$. The second derivative sign chart for p shows that $p''(x)$ is negative for $-1 < x < 2$ and positive for all other values of x . Since p'' changes sign at $x = -1$ and $x = 2$, p has points of inflection at $x = -1$ and $x = 2$. In particular, p is concave down for $-1 < x < 2$ and concave up for all other values of x .

- b. If we compute $p''\left(\frac{\sqrt{5}-1}{2}\right)$, we find that $p''\left(\frac{\sqrt{5}-1}{2}\right) \approx -1.205$, which means that p is concave down at $x = \frac{\sqrt{5}-1}{2}$. Since p has a horizontal tangent line at this point, it follows that p has a local maximum at $x = \frac{\sqrt{5}-1}{2}$ by the second derivative test.
- c. Since $p(2) = \frac{12}{e^2}$ and $p'(2) = -\frac{5}{e^2}$, the equation of the tangent line to p at $x = 2$ is given by $y = -\frac{5}{e^2}(x-2) + \frac{12}{e^2}$. To determine whether the tangent line lies above or below the graph of $y = p(x)$ at $x = 2$, we consider the concavity of p . Recall we know that $p''(2) = 0$ and that p is concave down to the left of $x = 2$ and concave up to the right. Thus, the tangent line lies neither solely below nor solely above the graph of $y = p(x)$ at $x = 2$; it lies below $y = p(x)$ to the right of $x = 2$ and above $y = p(x)$ to the left of $x = 2$.

3.2 Global Optimization

Motivating 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?

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. This information helps us determine the overall shape and behavior of the graph, as well as whether the function has relative extrema.

Remember the difference between a relative maximum and a global maximum: there is a *relative* maximum 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 3.2.1, 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 $x = b$.

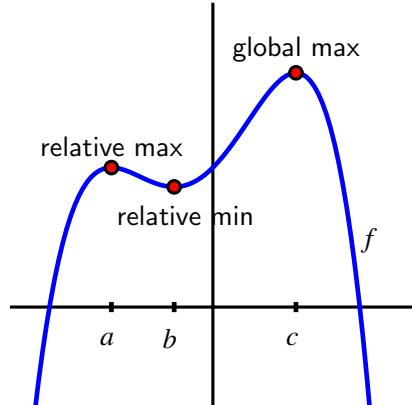


Figure 3.2.1: A function f with a global maximum, but no global minimum.

Our emphasis in this section is on finding the global extreme values of a function (if they exist), either over its entire domain or on some restricted portion.

Example 3.2.2. Let $f(x) = 2 + \frac{3}{1+(x+1)^2}$.

- Determine all of the critical numbers of f .
- Construct a first derivative sign chart for f and thus determine all intervals on which f is increasing or decreasing.
- Does f have a global maximum? If so, why, and what is its value and where is the maximum attained? If not, explain why.
- Determine $\lim_{x \rightarrow \infty} f(x)$ and $\lim_{x \rightarrow -\infty} f(x)$.
- Explain why $f(x) > 2$ for every value of x .
- Does f have a global minimum? If so, why, and what is its value and where is the minimum attained? If not, explain why.

3.2.1 Global Optimization

In Figure 3.2.1 and Example 3.2.2, we were interested in finding the global minimum and global maximum for f on its entire domain. At other times, we might focus on some restriction of the 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 numbers 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 example. 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.

Example 3.2.3. Let $g(x) = \frac{1}{3}x^3 - 2x + 2$.

- Find all critical numbers of g that lie in the interval $-2 \leq x \leq 3$.
- Use a graphing utility to construct the graph of g on the interval $-2 \leq x \leq 3$.
- From the graph, determine the x -values at which the absolute minimum and absolute maximum of g occur on the interval $[-2, 3]$.
- How do your answers change if we instead consider the interval $-2 \leq x \leq 2$?
- What if we instead consider the interval $-2 \leq x \leq 1$?

Hint.

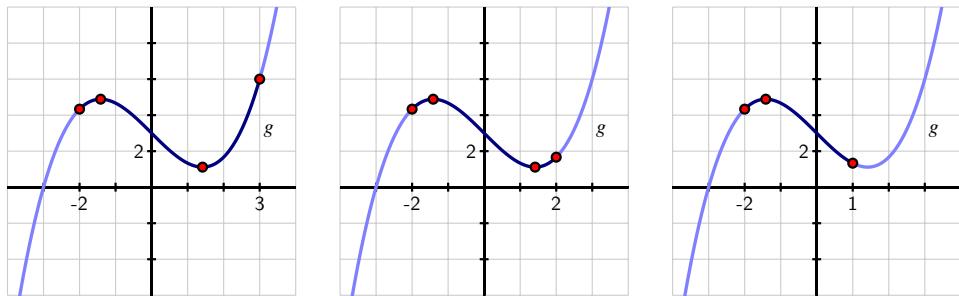
- Check that each critical number you find satisfies $-2 \leq x \leq 3$.
- <http://wolframalpha.com> is a great choice.

- c. On the graph, look for the lowest and highest possible values of the function.
- d. Ask yourself the same questions as (a)-(c), simply using the new interval.

Answer.

- a. $x = \pm\sqrt{2} \approx \pm 1.414$.

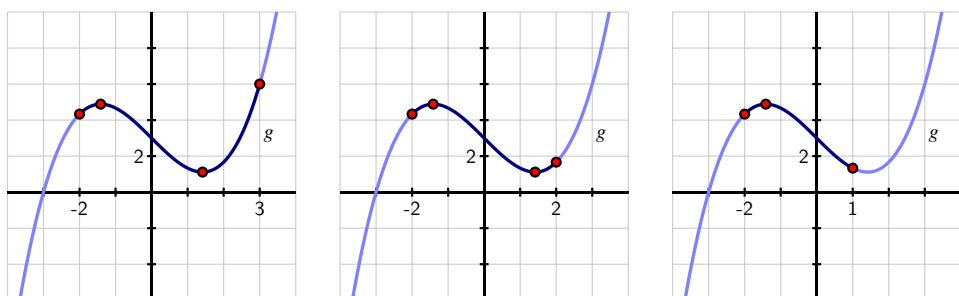
- b.



- c. On $[-2, 3]$, g has a global maximum at $x = 3$ and a global minimum at $x = \sqrt{2}$.
- d. On $[-2, 2]$, g has a global maximum at $x = -\sqrt{2}$ and a global minimum at $x = \sqrt{2}$.
- e. On $[-2, 3]$, g has a global maximum at $x = -\sqrt{2}$ and a global minimum at $x = 1$.

Solution.

- a. Since $g'(x) = x^2 - 2$, the critical numbers of g are $x = \pm\sqrt{2} \approx \pm 1.414$, both of which lie in the interval $-2 \leq x \leq 3$.
- b. The figure shown below shows three related plots, each with the emphases on the interval provided in (c), (d), and (e).



- c. On $[-2, 3]$, g has a global maximum at $x = 3$ and a global minimum at $x = \sqrt{2}$.
- d. On $[-2, 2]$, g has a global maximum at $x = -\sqrt{2}$ and a global minimum at $x = \sqrt{2}$.

- e. On $[-2, 3]$, g has a global maximum at $x = -\sqrt{2}$ and a global minimum at $x = 1$.

In Example 3.2.3, 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 numbers of the function, but also on the values of a and b . These observations demonstrate several important facts that hold 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 $f(x_M) \geq f(x)$ 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 on any closed interval $[a, b]$, a continuous function has to achieve both an absolute minimum and an absolute maximum. The theorem does not tell us where these extreme values occur, but rather only that they must exist. As we saw in Example 3.2.3, the only possible locations for relative extremes are at the endpoints of the interval or at a critical number.

Note 3.2.4. 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 numbers of f that lie in the interval;
- evaluate the function f at each critical number in the interval and at each endpoint of the interval;
- from among those function values, the smallest is the absolute minimum of f on the interval, while the largest is the absolute maximum.

Example 3.2.5. Find the *exact* absolute maximum and minimum of each function on the stated interval.

- $h(x) = xe^{-x}$, $[0, 3]$
- $p(t) = \sin(t) + \cos(t)$, $[-\frac{\pi}{2}, \frac{\pi}{2}]$
- $q(x) = \frac{x^2}{x-2}$, $[3, 7]$
- $f(x) = 4 - e^{-(x-2)^2}$, $(-\infty, \infty)$
- $h(x) = xe^{-ax}$, $[0, \frac{2}{a}]$ ($a > 0$)
- $f(x) = b - e^{-(x-a)^2}$, $(-\infty, \infty)$, $a, b > 0$

Hint.

- After computing $h'(x)$, factor to write the derivative as a product.

- b. The sine and cosine functions have the same value at $\frac{\pi}{4} \pm k\pi$ for any integer k .
- c. Upon finding $q'(x)$, factor its numerator.
- d. Remember that $e^{-(x-2)^2}$ is never zero.
- e. After differentiating, remove a factor of e^{-ax} .
- f. Compare part (d).

Answer.

- a. Absolute maximum: e^{-1} ; absolute minimum: 0.
- b. Absolute maximum: $\sqrt{2}$; absolute minimum: -1 .
- c. Absolute maximum: 9.8; absolute minimum: 8.
- d. Absolute minimum at $x = 2$; no absolute maximum.

Solution.

- a. For $h(x) = xe^{-x}$, we know that $h'(x) = xe^{-x}(-1) + e^{-x} = e^{-x}(-x + 1)$. Therefore, the only critical number of h is $x = 1$. Next, we compute $h(1)$, $h(0)$, and $h(3)$. Observe that
 - $h(1) = e^{-1} \approx 0.36788$
 - $h(0) = 0$
 - $h(3) = 3e^{-3} \approx 0.14936$

Thus, on $[0, 3]$, the absolute maximum of h is e^{-1} and the absolute minimum is 0.

- b. Given $p(t) = \sin(t) + \cos(t)$, it follows $p'(t) = \cos(t) - \sin(t)$, so $p'(t) = 0$ implies that $\cos(t) = \sin(t)$. The sine and cosine functions have the same value at $\frac{\pi}{4} \pm k\pi$ for any integer k . The only time this occurs in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is for $x = \frac{\pi}{4}$, and thus this is the only critical number of p in the given interval. Now,

- $p(\frac{\pi}{4}) = \sin(\frac{\pi}{4}) + \cos(\frac{\pi}{4}) = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = \sqrt{2} \approx 1.41421$
- $p(-\frac{\pi}{2}) = \sin(-\frac{\pi}{2}) + \cos(-\frac{\pi}{2}) = -1 + 0 = -1$
- $p(\frac{\pi}{2}) = \sin(\frac{\pi}{2}) + \cos(\frac{\pi}{2}) = 1 + 0 = 1$

Therefore, on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, the absolute maximum of p is $\sqrt{2}$ and the absolute minimum is -1 .

- c. With $q(x) = \frac{x^2}{x-2}$, we have

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

Hence, the critical numbers of q are $x = 0$ and $x = 4$. Only the latter critical number lies in the interval $[3, 7]$, and thus we evaluate q and find

- $q(4) = \frac{16}{2} = 8$
- $q(3) = \frac{9}{1} = 9$
- $q(7) = \frac{49}{5} = 9.8$

We now see that on $[3, 7]$ the absolute maximum of q is 9.8 and the absolute minimum is 8.

- d. Here, we first observe that we are working on the domain of all real numbers, not a closed bounded interval. Hence, we need to think about the overall behavior of the function. First, since $f(x) = 4 - e^{-(x-2)^2}$, by the chain rule we see that $f'(x) = -e^{-(x-2)^2}(-2(x-2)) = 2(x-2)e^{-(x-2)^2}$. Since $e^{-(x-2)^2}$ is always positive (in particular, never zero), it follows that the only critical number of f is $x = 2$. Furthermore, with $f'(x) = 2(x-2)e^{-(x-2)^2}$, we see that for $x < 2$, $f'(x) < 0$, while for $x > 2$, $f'(x) > 0$. This tells us by the first derivative test that f is decreasing for $x < 2$ and increasing for $x > 2$, which tells us that f has an absolute minimum at $x = 2$, and f does not have an absolute maximum.

The interval we choose has nearly the same influence on extreme values as the function under consideration. Consider, for instance, the function pictured in Figure 3.2.6.

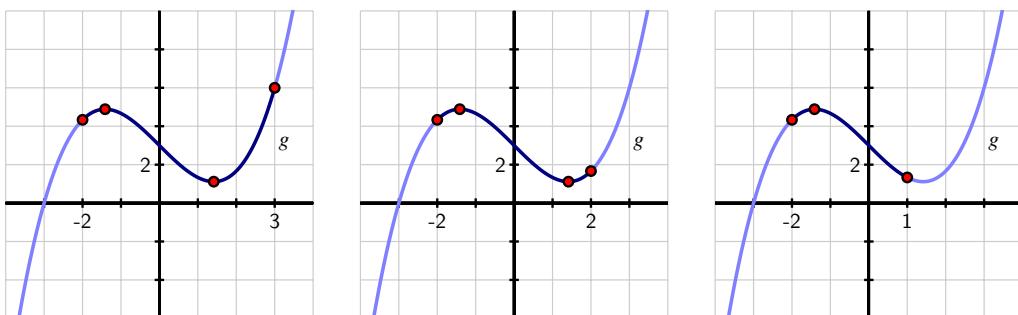


Figure 3.2.6: A function g considered on three different intervals.

In sequence, from left to right, the interval under consideration is changed from $[-2, 3]$ to $[-2, 2]$ to $[-2, 1]$.

- On the interval $[-2, 3]$, there are two critical numbers, with the absolute minimum at one critical number and the absolute maximum at the right endpoint.
- On the interval $[-2, 2]$, both critical numbers are in the interval, with the absolute minimum and maximum at the two critical numbers.
- On the interval $[-2, 1]$, just one critical number lies in the interval, with the absolute maximum at one critical number and the absolute minimum at one endpoint.

Remember to consider only the critical numbers that lie within the interval.

3.2.2 Moving toward applications

We conclude this section with an example of an applied optimization problem. It highlights the role that a closed, bounded domain can play in finding absolute extrema.

Example 3.2.7. A 20 cm piece of wire is cut into two pieces. One piece is used to form a square and the other to form an equilateral triangle. How should the wire be cut to maximize the total area enclosed by the square and triangle? to minimize the area?

Solution. We begin by sketching a picture that illustrates the situation. The variable in the problem is where we decide to cut the wire. We thus label the cut point at a distance x from one end of the wire, and note that the remaining portion of the wire then has length $20 - x$.

As shown in Figure 3.2.8, we see that the x cm of wire that is used to form the equilateral 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 $\frac{20-x}{4}$.

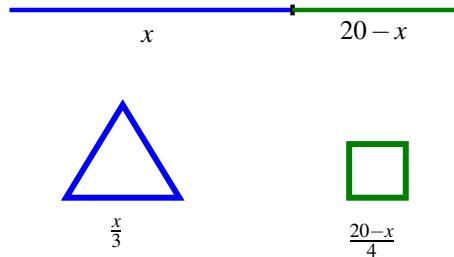


Figure 3.2.8: A 20 cm piece of wire cut into two pieces, one of which forms an equilateral triangle, the other which yields a square.

At this point, we note that there are obvious restrictions on x : in particular, $0 \leq x \leq 20$. 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 minimum and maximum areas that can be enclosed. Because the height of an equilateral triangle is $\sqrt{3}$ times half the length of the base, the area of the triangle is

$$A_{\Delta} = \frac{1}{2}bh = \frac{1}{2} \cdot \frac{x}{3} \cdot \frac{x\sqrt{3}}{6}.$$

The area of the square is $A_{\square} = s^2 = \left(\frac{20-x}{4}\right)^2$. Therefore, the total area function is

$$A(x) = \frac{\sqrt{3}x^2}{36} + \left(\frac{20-x}{4}\right)^2.$$

Remember that we are considering this function only on the restricted domain $[0, 20]$.

Differentiating $A(x)$, we have

$$A'(x) = \frac{\sqrt{3}x}{18} + 2\left(\frac{20-x}{4}\right)\left(-\frac{1}{4}\right) = \frac{\sqrt{3}}{18}x + \frac{1}{8}x - \frac{5}{2}.$$

When we set $A'(x) = 0$, we find that $x = \frac{180}{4\sqrt{3}+9} \approx 11.3007$ is the only critical number of A in the interval $[0, 20]$.

Evaluating A at the critical number 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. 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 3.2.9.

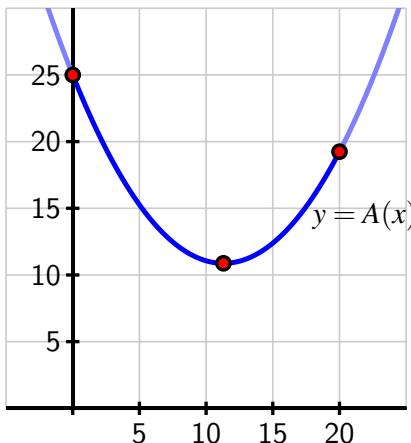


Figure 3.2.9: A plot of the area function from Example 3.2.7.

Example 3.2.10. 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.

- 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.
- Determine a formula for the function V (that depends on the variable in (a)) that tells us the volume of the box.

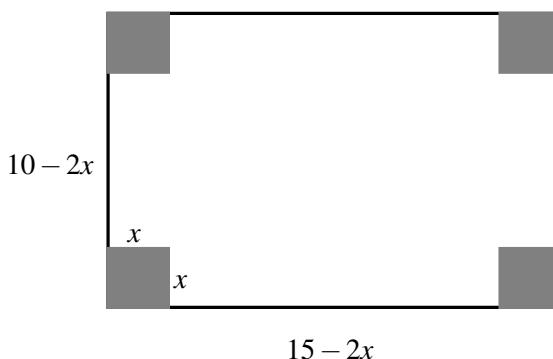
- c. What is the domain of the function V ? That is, what values of x make sense for input? Are there additional restrictions provided in the problem?
- d. Determine all critical numbers of the function V .
- e. Evaluate V at each of the endpoints of the domain and at any critical numbers that lie in the domain.
- f. What is the maximum possible volume of the box? the minimum?

Hint.

- a. Consider letting the length of one side of the removed squares be represented by x .
- b. Remember that the volume of a box is length \times width \times height.
- c. Read the given information carefully and think about the picture.
- d. Note that since V is a cubic function, V' is quadratic.
- e. Which critical numbers satisfy $1 \leq x \leq 3$?
- f. Evaluate the function at appropriate points.

Answer.

a.



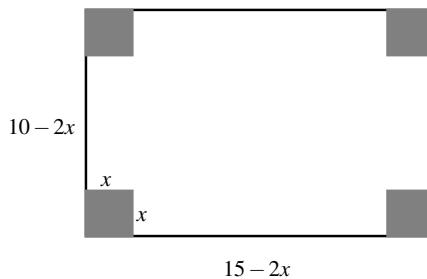
- b. $V(x) = x(10 - 2x)(15 - 2x) = 4x^3 - 50x^2 + 150x$.
- c. $1 \leq x \leq 3$.
- d. $x = \frac{25 \pm 5\sqrt{7}}{6} \approx 6.371459426, 1.961873908$.
- e. • $V(1.961873908) = 132.0382370$

- $V(1) = 104$
- $V(3) = 108$

f. Absolute maximum: 132.0382370; absolute minimum: 104.

Solution.

- a. We let x represent the length of a side of the square that is cut from each corner, so that we have the following picture:



- b. Because the box has dimensions $(10 - 2x) \times (15 - 2x) \times x$, the volume of the box is given by

$$V(x) = x(10 - 2x)(15 - 2x) = 4x^3 - 50x^2 + 150x.$$

- c. Clearly the smallest x can be is 0 and the largest x can be is 5, since one side of the cardboard has length 10. But we're told in the problem to restrict the value of x to $1 \leq x \leq 3$, so this is the domain we use for V , even though V is defined for every real number x .

- d. Since $V'(x) = 12x^2 - 100x + 150$, it follows that the critical numbers (where $V'(x) = 0$) are

$$x = \frac{25 \pm 5\sqrt{7}}{6} \approx 6.371459426, 1.961873908.$$

- e. Only the latter critical number is in the relevant domain of V , and hence we consider

- $V(1.961873908) = 132.0382370$
- $V(1) = 104$
- $V(3) = 108$

- f. Hence the absolute maximum possible volume of the box is 132.0382370 and occurs when $x = 1.961873908$, while the absolute minimum is 104, which occurs when $x = 1$.

Example 3.2.7 and Example 3.2.10 illustrate 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, find a function that models

the quantity to be optimized, and then decide on an appropriate domain for that function. Once that is done, we are in the familiar situation of finding the absolute minimum and maximum of a function over a particular domain, so we apply the calculus ideas that we have been studying to this point in Chapter 3.

3.2.3 Summary

- To find relative extreme values of a function, we use a first derivative sign chart and classify all of the function's critical numbers. 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. If we are working to find absolute extremes on a restricted interval, then we first identify all critical numbers 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 numbers and the endpoints. Thus, we simply evaluate the function at each endpoint and each critical number in the interval, and compare the results to decide which is largest (the absolute maximum) and which is smallest (the absolute minimum).

3.2.4 Exercises

1. 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.

- f is a function such that $f''(x) < 0$ for every x .
- g is a function with two critical numbers 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 $x > b$.
- h is a function with two critical numbers 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, $\lim_{x \rightarrow \infty} h(x) = 0$ and $\lim_{x \rightarrow -\infty} h(x) = 0$.
- p is a function differentiable everywhere except at $x = a$ and $p''(x) > 0$ for $x < a$ and $p''(x) < 0$ for $x > a$.

Answer.

- Not enough information is given.
- Global minimum at $x = a$.
- Global minimum at $x = a$; global maximum at $x = b$.
- Not enough information is provided.

Solution.

- a. It's not possible to say without more information: for instance, the functions $f(x) = e^{-x}$, $f(x) = e^{-x}$, and $f(x) = x^2$ are each always concave up, but only one of them has a global minimum. We can say for sure that by being always concave up, f cannot have a global maximum, and it also cannot have both a global maximum and a global minimum. It might have a global minimum, or it might have neither.
- b. By being twice differentiable, we know that $g'(x) = 0$ any time x is a critical number of g . Thus, given the signs of $g'(x)$ on the stated intervals, g is decreasing before $x = a$ and decreasing on $a < x < b$, with a horizontal tangent line at $x = a$. In addition, g is increasing for $x > b$. Thus, we know that g has a global minimum at $x = a$.
- c. Since h is always decreasing for $x < a$, always increasing for $a < x < b$, and always decreasing for $x > b$, we know that h has a local minimum at $x = a$ and a local maximum at $x = b$. Moreover, these local extremes are both global extremes since $h(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. In particular, since $h(x)$ decreases to 0 for $x > b$ and increases on $a < x < b$, it follows that $h(b)$ is the largest value of $h(x)$ on (a, ∞) , and that $h(b) > 0$. Similarly, since $h(x)$ decreases on $(-\infty, a)$ and approaches 0 as $x \rightarrow -\infty$, and $h(x)$ increases on $a < x < b$, it must be the case that $h(a)$ is the least value of $h(x)$ on $(-\infty, b)$ and that $h(b) < 0$. Hence, $h(x)$ is always negative on $x < a$ and $h(x)$ is always positive on $x > b$. From all of these observations, it follows that h has a global minimum at $x = a$ and a global maximum at $x = b$.
- d. First, observe that p is defined for all real numbers by assumption, so $p(a)$ has some finite value. However, p is not differentiable at $x = a$ and p is concave up to the left of $x = a$ and concave down to the right of $x = a$. This indicates that p has a relative maximum at $x = a$. Not enough information is provided, however, to determine if this relative maximum is a global maximum, nor if p happens to have a global minimum at some other point.
2. For each family of functions that depends on one or more parameters, determine the function's absolute maximum and absolute minimum on the given interval.
- $p(x) = x^3 - a^2x$, $[0, a]$ ($a > 0$)
 - $r(x) = axe^{-bx}$, $[\frac{1}{2b}, \frac{2}{b}]$ ($a > 0, b > 1$)
 - $w(x) = a(1 - e^{-bx})$, $[b, 3b]$ ($a, b > 0$)
 - $s(x) = \sin(kx)$, $[\frac{\pi}{3k}, \frac{5\pi}{6k}]$ ($k > 0$)

Answer.

- Global maximum $p(0) = p(a) = 0$; global minimum $p\left(\frac{a}{\sqrt{3}}\right) = -\frac{2a^3}{3\sqrt{3}}$.
- Global max $r\left(\frac{1}{b}\right) \approx 0.368\frac{a}{b}$; global min $r\left(\frac{2}{b}\right) \approx 0.270\frac{a}{b}$.
- Global minimum $g(b) = a(1 - e^{-b^2})$; global maximum $g(3b) = a(1 - e^{-3b^2})$.
- Global max $s\left(\frac{\pi}{2k}\right) = 1$; global min $s\left(\frac{5\pi}{6k}\right) = \frac{1}{2}$.

Solution.

- Given $p(x) = x^3 - a^2x$ on $[0, a]$, we first find the critical number(s) of p that lie in $[0, a]$. Since $p'(x) = 3x^2 - a^2 = 3\left(x^2 - \frac{a^2}{3}\right)$, we know p has critical numbers at $x = \pm\frac{a}{\sqrt{3}}$. The negative critical number lies outside the interval, while the positive one lies within the

interval since $0 < \frac{1}{\sqrt{3}} < 1$. Evaluating p at the endpoints and the critical number within the interval, we find that $p(0) = 0$, $p(a) = a^3 - a^2 \cdot a = 0$, and $p\left(\frac{a}{\sqrt{3}}\right) = \left(\frac{a}{\sqrt{3}}\right)^3 - a^2 \cdot \left(\frac{a}{\sqrt{3}}\right) = \frac{a^3}{3\sqrt{3}} - \frac{a^3}{\sqrt{3}} = -\frac{2a^3}{3\sqrt{3}}$. Hence, the global maximum of p on $[0, a]$ is $p(0) = p(a) = 0$ and the global minimum is $p\left(\frac{a}{\sqrt{3}}\right) = -\frac{2a^3}{3\sqrt{3}}$.

- b. Given $r(x) = axe^{-bx}$ on $[\frac{1}{2b}, \frac{2}{b}]$, we first find the critical numbers of r that lie in the given interval. Note that by the product rule, $r'(x) = axe^{-bx}(-b) + e^{-bx}a$. Factoring, $r'(x) = ae^{-bx}(-bx + 1)$, and thus the only critical number of r occurs where $x = \frac{1}{b}$. Note that $\frac{1}{2b} < \frac{1}{b} < b$ since $b > 1$, and thus the critical number lies within the interval.

Evaluating r at the endpoints and the critical number, we find

- $r\left(\frac{1}{2b}\right) = a \cdot \frac{1}{2b} \cdot e^{-1/2} \approx 0.607 \frac{a}{2b} = 0.3035 \frac{a}{b}$
- $r\left(\frac{2}{b}\right) = a \cdot \frac{2}{b} \cdot e^{-2} \approx 0.135 \frac{2a}{b} = 0.270 \frac{a}{b}$
- $r\left(\frac{1}{b}\right) = a \cdot \frac{1}{b} \cdot e^{-1} \approx 0.368 \frac{a}{b}$

The global max of r is therefore $r\left(\frac{1}{b}\right) \approx 0.368 \frac{a}{b}$ and the global min is $r\left(\frac{2}{b}\right) \approx 0.270 \frac{a}{b}$.

- c. For $w(x) = a(1 - e^{-bx})$ on $[b, 3b]$, we first note that $w'(x) = a(be^{-bx}) = abe^{-bx}$. Since a and b are both positive and so is e^{-bx} for every value of x , we see that $w'(x) > 0$ for every value of x . Because w is always increasing (specifically on $[b, 3b]$), w has its global minimum at the left endpoint ($w(b) = a(1 - e^{-b^2})$) and its global maximum at the right endpoint ($w(3b) = a(1 - e^{-3b^2})$).
- d. Finally, for $s(x) = \sin(kx)$ on $[\frac{\pi}{3k}, \frac{5\pi}{6k}]$, we know $s'(x) = k \cos(kx)$, and $\cos(kx) = 0$ when $kx = \frac{\pi}{2} \pm n\pi$ for some integer n , so $x = \frac{\pi}{2k} \pm n\pi = \frac{\pi \pm 2n\pi}{2k}$. These x -values are the critical numbers of s . Note that when $n = -1, 0, 1$, the corresponding critical numbers are

$$x = -\frac{\pi}{2k}, \frac{\pi}{2k}, \frac{3\pi}{2k}$$

Comparing the values of $-\frac{\pi}{2k}$, $\frac{\pi}{2k}$, and $\frac{3\pi}{2k}$ to $\frac{\pi}{3}$ and $\frac{5\pi}{6}$, we see that only $\frac{\pi}{2k}$ lies in the given interval $[\frac{\pi}{3k}, \frac{5\pi}{6k}]$, and thus $x = \frac{\pi}{2k}$ is the only critical number at which we need to evaluate s .

Evaluating s at the endpoints and relevant critical number, we find:

- $s\left(\frac{\pi}{3k}\right) = \sin\left(k \cdot \frac{\pi}{3k}\right) = \sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}$;
- $s\left(\frac{5\pi}{6k}\right) = \sin\left(k \cdot \frac{5\pi}{6k}\right) = \sin\left(\frac{5\pi}{6}\right) = \frac{1}{2}$; and
- $s\left(\frac{\pi}{2k}\right) = \sin\left(k \cdot \frac{\pi}{2k}\right) = \sin\left(\frac{\pi}{2}\right) = 1$.

Hence the global max of s on $[\frac{\pi}{3k}, \frac{5\pi}{6k}]$ is $s\left(\frac{\pi}{2k}\right) = 1$ and the global min of s is $s\left(\frac{5\pi}{6k}\right) = \frac{1}{2}$.

3. 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 numbers 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. $f'(x) \leq 0$ for all x in $[a, b]$
- b. g has a critical number 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 number c such that $a < c < b, p'(x) < 0$ for $x < c$ and $p'(x) > 0$ for $x > c$

Answer.

- a. Global maximum at $x = a$; global minimum at $x = b$.
- b. Global maximum at $x = c$; global minimum at either $x = a$ or $x = b$.
- c. Global minimum at $x = a$ and $x = b$; global maximum somewhere in (a, b) .
- d. Global minimum at $x = c$; global maximum value at $x = a$.

Solution.

- a. The fact that $f'(x) \leq 0$ for all x in $[a, b]$ means that f cannot increase in this interval. So either f is constant or f decreases somewhere in this interval. We can conclude that the global maximum value of f occurs at $x = a$ and the global minimum value at $x = b$.
 - b. The information we are given shows that g is increasing on (a, c) and decreasing on (c, b) . So g has a global maximum value at $x = c$. We know that g must have a global minimum at either $x = a$ or $x = b$. The information given does not tell us how much g increases on (a, c) and decreases on (c, b) , so we cannot determine the endpoint at which g attains its global minimum value.
 - c. In order for $h(a) = h(b)$ and $h''(x) < 0$ for all x in $[a, b]$, the graph of h must look something like a parabola that opens down on the interval $[a, b]$. Consequently, h has its global minimum at $x = a$ and $x = b$, and its global maximum somewhere in (a, b) , but we do not know the exact point at which h attains its global maximum.
 - d. The given information forces the graph of p to decrease on (a, c) then increase on (c, b) . Thus, we must have $p(c) < p(b)$ and so p has its global minimum value at $x = c$. Since $p(a) > p(b)$, this makes $p(a)$ the global maximum value of p on $[a, b]$.
4. 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 numbers of s .

- a. $[\frac{\pi}{6}, \frac{7\pi}{6}]$
- b. $[0, \frac{\pi}{2}]$
- c. $[0, 2\pi]$
- d. $[\frac{\pi}{3}, \frac{5\pi}{6}]$

Answer.

- a. Global max $s(\frac{5\pi}{12}) = 8$; global min $s(\frac{11\pi}{12}) = 2$.
- b. Global max $s(\frac{5\pi}{12}) = 8$; global min $s(0) = 5 - \frac{3\sqrt{3}}{2} \approx 2.402$.
- c. Global max $s(\frac{5\pi}{12}) = 8$; global min $s(\frac{11\pi}{12}) = 2$. (There are other points at which the function achieves these values on the given interval.)
- d. Global max $s(\frac{5\pi}{12}) = 8$; global min $s(\frac{\pi}{3}) = 6.5$.

Solution.

- a. For $s(t) = 3 \sin(2(t - \frac{\pi}{6})) + 5$, $s'(t) = 3 \cos(2(t - \frac{\pi}{6})) \cdot 2 = 6 \cos(2(t - \frac{\pi}{6}))$. It follows that $s'(t) = 0$ if and only if $\cos(2(t - \frac{\pi}{6})) = 0$, and thus $2(t - \frac{\pi}{6}) = \frac{\pi}{2} \pm k\pi$ for some integer value of k . Solving this most recent equation for t , we have $t - \frac{\pi}{6} = \frac{\pi}{4} \pm \frac{k\pi}{2}$, so

$$t = \frac{\pi}{6} + \frac{\pi}{4} \pm \frac{k\pi}{2} = \frac{5\pi}{12} \pm \frac{k\pi}{2}.$$

Now we consider the given interval $[\frac{\pi}{6}, \frac{7\pi}{6}]$, which we can think of as $[\frac{2\pi}{12}, \frac{14\pi}{12}]$. The value $t = \frac{5\pi}{12}$ lies in this interval, as does $t = \frac{5\pi}{12} + \frac{\pi}{2} = \frac{11\pi}{12}$. Thus, we have two critical numbers in the interval at which to evaluate s , as well as at the endpoints.

Doing so, we find:

- $s(\frac{5\pi}{12}) = 8$;
- $s(\frac{11\pi}{12}) = 2$;
- $s(\frac{\pi}{6}) = 5$; and
- $s(\frac{7\pi}{6}) = 5$.

The global max of s on $[\frac{\pi}{6}, \frac{7\pi}{6}]$ is thus $s(\frac{5\pi}{12}) = 8$, while the global min is $s(\frac{11\pi}{12}) = 2$.

- b. From our work in (a), we know that s has critical numbers at $t = \frac{5\pi}{12} \pm \frac{k\pi}{2}$. On the interval $[0, \frac{\pi}{2}]$, only $t = \frac{5\pi}{12}$ lies in the interval. Thus we have three values at which to evaluate s .

Doing so:

- $s(\frac{5\pi}{12}) = 8$;
- $s(0) = 3 \sin(2(0 - \frac{\pi}{6})) + 5 = 3 \sin(-\frac{\pi}{3}) + 5 = 5 - \frac{3\sqrt{3}}{2} \approx 2.402$; and
- $s(\frac{\pi}{2}) = 3 \sin(2(\frac{\pi}{2} - \frac{\pi}{6})) + 5 = 3 \sin(\frac{2\pi}{3}) + 5 = 5 + \frac{3\sqrt{3}}{2} \approx 7.598$.

The global max of s on $[0, \frac{\pi}{2}]$ is thus $s(\frac{5\pi}{12}) = 8$, while the global min is $s(0) = 5 - \frac{3\sqrt{3}}{2} \approx 2.402$.

- c. On the interval from $t = 0$ to $t = 2\pi$, s completes nearly two full periods, and thus achieves each of its maximum and minimum values. The global maximum is 8 as determined in (a), and the global minimum is 2. These occur, for example, at $s(\frac{5\pi}{12}) = 8$, while the global min is $s(\frac{11\pi}{12}) = 2$. There are other points at which the function achieves these values on the given interval.

- d. On the interval $[\frac{\pi}{3}, \frac{5\pi}{6}] = [\frac{4\pi}{12}, \frac{10\pi}{12}]$, only the critical number $t = \frac{5\pi}{12}$ lies in the interval. Thus we have three values at which to evaluate s .

Doing so:

- $s(\frac{5\pi}{12}) = 8$;
- $s(\frac{\pi}{3}) = 3 \sin(2(\frac{\pi}{3} - \frac{\pi}{6})) + 5 = 3 \sin(\frac{\pi}{6}) + 5 = \frac{3}{2} + 5 = \frac{13}{2} = 6.5$; and
- $s(\frac{5\pi}{6}) = 3 \sin(2(\frac{5\pi}{6} - \frac{\pi}{6})) + 5 = 3 \sin(\frac{2\pi}{3}) + 5 = 5 + \frac{3\sqrt{3}}{2} \approx 7.598$.

The global max of s on $[\frac{\pi}{3}, \frac{5\pi}{6}]$ is thus $s(\frac{5\pi}{12}) = 8$, while the global min is $s(\frac{\pi}{3}) = 6.5$.

3.3 Applied Optimization

Motivating 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?
-

Near the conclusion of Section 3.2, we considered two optimization problems where determining the function to be optimized was part of the problem. In Example 3.2.7, we sought to use a single piece of wire to build an equilateral triangle and square in order to maximize the total combined area enclosed. In the subsequent Example 3.2.10, 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.

In neither of these problems was a function to optimize explicitly provided. Rather, we first tried to understand the problem by drawing a figure and introducing variables, and then sought to develop a formula for a function that modeled the quantity to be optimized. Once the function was established, we then considered what domain was appropriate. At that point, we were finally ready to apply the ideas of calculus to determine the absolute minimum or maximum.

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.

Example 3.3.1. 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 3.3.2.

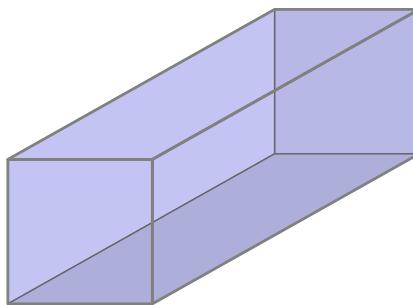


Figure 3.3.2: A rectangular parcel with a square end.

- 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.

3.3.1 More applied optimization problems

Many of the steps in Example 3.3.1 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.

Note 3.3.3.

- 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 Example 3.3.1, 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 numbers 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 numbers (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 3.3.4. 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.

Hint.

- a. Note that both the radius and the height of the can are variable.

- b. Remember that volume is the area of the base times the height, while surface area can be thought of in terms of the area of the two lids, plus the area of the “side” of the can.
- c. Use the fact that $V = 16$ to write one of the variables in terms of the other to get the cost as a function of a single variable.
- d. Differentiate the total cost function and find its critical number(s) first.

Answer.

- a. Let the can have radius r and height h .
- b. $V = \pi r^2 h$; $S = 2\pi r^2 + 2\pi r h$; $C = 2\pi r^2 \cdot 0.027 + 2\pi r h \cdot 0.015$.
- c. $C(r) = 0.054\pi r^2 + 0.48\frac{1}{r}$, $r > 0$.
- d. $r = \sqrt[3]{\frac{0.48}{0.108\pi}} \approx 1.12259$; $h \approx 4.041337$; minimum cost $C(1.12259) \approx 0.64137$.

Solution.

- a. We let r be the radius of the base of the cylindrical can and h be its height.
- b. Volume is the area of the base times the height, so $V = \pi r^2 h$. Surface area is the area of the lids plus the area of the side, the latter of which is a rectangle with height h and width the perimeter of the base. Hence, $S = 2\pi r^2 + 2\pi r h$. Finally, the total cost is the cost of the lids plus the cost of the sides, which is

$$C = 2\pi r^2 \cdot 0.027 + 2\pi r h \cdot 0.015.$$

- c. Because the volume is fixed at 16 cubic inches, we know that $16 = \pi r^2 h$. Solving for h , $h = \frac{16}{\pi r^2}$. Substituting this expression for h in the formula for total cost, we now have that

$$C = 2\pi r^2 \cdot 0.027 + 2\pi r \left(\frac{16}{\pi r^2} \right) \cdot 0.015 = 0.054\pi r^2 + 0.48\frac{1}{r}.$$

With $C(r) = 0.054\pi r^2 + 0.48\frac{1}{r}$, we note that the only constraint on r is that $r > 0$, hence this is the domain on which we seek to minimize C .

- d. Noting that $C'(r) = 0.108\pi r - 0.48\frac{1}{r^2}$, we set $C'(r) = 0$ and solve for r to find that

$$0.108\pi r = \frac{0.48}{r^2},$$

so that $r^3 = \frac{0.48}{0.108\pi} \approx 1.41471$, from which it follows that $r = \sqrt[3]{\frac{0.48}{0.108\pi}} \approx 1.12259$ is the only critical number of C . At this point, we can use either the first or second derivative test to justify that C has an absolute minimum at $r = 1.12259$. We choose to use the second derivative test; note that $C''(r) = 0.108\pi + 0.96\frac{1}{r^3}$, which is always positive for $r > 0$, hence C is always concave up on the relevant domain ($r > 0$), which makes $r = \sqrt[3]{\frac{0.48}{0.108\pi}} \approx 1.12259$ where the absolute minimum of C occurs. In addition, we note that since $h = \frac{16}{\pi r^2}$, the corresponding h value is $h \approx 4.041337$, and the overall minimum cost is $C(1.12259) \approx 0.64137$.

Familiarity with common geometric formulas is particularly helpful in problems such as the one in Example 3.3.4. 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 the variables.

Example 3.3.5. 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 3.3.6. In order to minimize the time to go from P to Z to the cabin, where should the hiker turn into the forest?

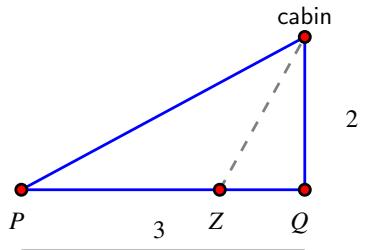


Figure 3.3.6: A hiker walks from P to Z to the cabin, as pictured.

Hint. Let x be the distance from Z to Q . What is then the distance from P to Z in terms of x ? How about the distance from Z to the cabin? How does time depend on distance and rate?

Answer. The absolute minimum time the hiker can achieve is 0.99302 hours, which is attained by hiking about 2.2 km from P to Q and then turning into the woods for the remainder of the trip.

Solution. We begin by letting x be the distance from Z to Q . Since it is 3 km from P to Q , the distance from P to Z is $3 - x$. Further, by the Pythagorean Theorem, the distance from Z to the cabin is $\sqrt{4 + x^2}$.

Next, we want to determine the hiker's time as a function of x . Because distance equals rate times time, time is thus distance divided by rate. Along the road, the hiker's distance is $3 - x$ km and her rate is 8 km/hr, thus her time on the road, T_r is

$$T_r = \frac{3 - x}{8}.$$

Once she enters the woods, her rate drops to 3 km/hr and travels a distance of $\sqrt{4 + x^2}$ km,

making her time in the woods, T_w , given by

$$T_w = \frac{\sqrt{4+x^2}}{3}.$$

Thus, the hiker's total time is given by the function

$$T(x) = \frac{3-x}{8} + \frac{\sqrt{4+x^2}}{3}.$$

Because the only values of x that make sense to use are $0 \leq x \leq 3$ (using either negative values or values greater than 3 clearly add unnecessary time to the trip), we use this domain for T and now seek the absolute minimum of T on $[0, 3]$. First, we find that

$$T'(x) = -\frac{1}{8} + \frac{1}{3} \cdot \frac{1}{2}(4+x^2)^{-1/2}(2x) = -\frac{1}{8} + \frac{x}{3\sqrt{4+x^2}}.$$

Setting $T'(x) = 0$ and solving for x , we have $\frac{x}{3\sqrt{4+x^2}} = \frac{1}{8}$, so $8x = 3\sqrt{4+x^2}$. Squaring both sides, $64x^2 = 9(4+x^2) = 36 + 9x^2$. Hence, $55x^2 = 36$, so $x = \sqrt{\frac{36}{55}} \approx 0.80904$. (We don't consider the critical number $x = -\sqrt{\frac{36}{55}}$ because this doesn't lie in the relevant domain of T .)

Finally, we evaluate T at the only critical number in the interval and at the interval's endpoints. Doing so, we find $T(0) = \frac{3}{8} + \frac{2}{3} \approx 1.0417$, $T(3) = \frac{\sqrt{13}}{3} \approx 1.20185$, and $T(\sqrt{\frac{36}{55}}) = \frac{3}{8} + \frac{\sqrt{55}}{12} \approx 0.99302$, and thus the absolute minimum time the hiker can achieve is 0.99302 hours, which is attained by hiking about 2.2 km from P to Q and then turning into the woods for the remainder of the trip.

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 example is similar to the latter problem.

Example 3.3.7. 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 $f(x) = b - ax^2$.)

Hint. Let x represent half the width of the rectangle's base. How does the rectangle's height depend on x ?

Answer. Maximum area: $A(\frac{5}{\sqrt{3}}) = \frac{500}{9}\sqrt{3} \approx 96.225$. Maximum perimeter: $P(1) = 52$. At $x = \frac{\sqrt{82}-1}{3}$ the absolute maximum of combined perimeter and area occurs.

Solution. Letting x represent half the width of the rectangle's base, it follows that the rectangle's height is $25 - x^2$. Hence the area of the rectangle is

$$A(x) = 2x(25 - x^2) = 50x - 2x^3.$$

Based on the region, the only possible values of x are for $0 \leq x \leq 5$; moreover, it is evident that for either $x = 0$ or $x = 5$, the area of the corresponding rectangle is zero, which can't be where the maximum occurs. Differentiating,

$$A'(x) = 50 - 6x^2,$$

so we find the critical number of A by solving $6x^2 = 50$, which yields $x = \frac{5}{\sqrt{3}} \approx 2.8868$, which results in the maximum possible area of

$$A\left(\frac{5}{\sqrt{3}}\right) = \frac{500}{9}\sqrt{3} \approx 96.225$$

square units.

To maximize perimeter, we note that the rectangle's perimeter is

$$P(x) = 2x + (25 - x^2) + 2x + (25 - x^2) = -2x^2 + 4x + 50.$$

It is straightforward to show that the only critical number of the quadratic function P occurs when $x = 1$ and that the corresponding absolute maximum value is $P(1) = 52$.

Finally, to consider combined area and perimeter, examine the function $f(x) = A(x) + P(x)$ on the interval $0 \leq x \leq 5$. You should find that the only relevant critical number is $x = \frac{\sqrt{82}-1}{3}$ and that the absolute maximum of f occurs at that value.

Example 3.3.8. 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 3.3.9. At what angle should the folds be made to produce the trough of maximum volume?

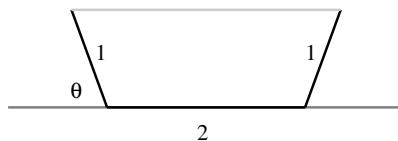


Figure 3.3.9: A cross-section of the trough formed by folding to an angle of θ .

Hint. Drop altitudes from the top of the trough to the base of length 2. In the two triangles that are formed, what are the lengths of the legs in terms of θ ?

Answer. $A(1.19606) \approx 2.2018$ is the absolute maximum cross-sectional area, which leads to the absolute maximum volume.

Solution. Once we choose the angle θ , the two right triangles in the trapezoid are determined, and each has a horizontal leg of length $\cos(\theta)$ and a vertical leg of length $\sin(\theta)$. Thus, the sum of the areas of the two triangles is $\sin(\theta)\cos(\theta)$, and the area of the rectangle between them is $2\sin(\theta)$. Hence, the area of the trapezoidal cross-section is $A(\theta) = \sin(\theta)\cos(\theta) + 2\sin(\theta)$. Because the length of the trough is constant, the trough's volume will be maximized by maximizing cross-sectional area. Note, too, that the domain for θ is $0 \leq \theta \leq \frac{\pi}{2}$.

Differentiating, we find that

$$A'(\theta) = \sin(\theta)(-\sin(\theta)) + \cos(\theta)\cos(\theta) + 2\cos(\theta) = -\sin^2(\theta) + \cos^2(\theta) + 2\cos(\theta).$$

Using the identity $\sin^2(\theta) = 1 - \cos^2(\theta)$, it follows that

$$A'(\theta) = \cos^2(\theta) - 1 + \cos^2(\theta) + 2\cos(\theta) = 2\cos^2(\theta) + 2\cos(\theta) - 1.$$

This most recent equation is quadratic in $\cos(\theta)$, so letting $u = \cos(\theta)$, we can start to solve the equation $A'(\theta) = 0$ by solving $2u^2 + 2u - 1 = 0$. Doing so, we find that

$$u = \frac{-1 \pm \sqrt{3}}{2} \approx 0.3660254, -1.3660254,$$

so only $u = \frac{-1+\sqrt{3}}{2}$ will be a potential value of the cosine function from an angle that lies in the interval $[0, \frac{\pi}{2}]$. Now, recalling that $\cos(\theta) = u$, to find the critical number θ , we solve $\cos(\theta) = \frac{-1+\sqrt{3}}{2}$, which implies $\theta = \arccos(\frac{-1+\sqrt{3}}{2}) \approx 1.19606$ radians, or $\theta \approx 68.5292^\circ$.

Finally, to confirm that A has an absolute maximum at $\theta = \arccos(\frac{-1+\sqrt{3}}{2}) \approx 1.19606$, we consider this value as well as the endpoints of $[0, \frac{\pi}{2}]$, and evaluate A to find that $A(0) = 0$, $A(\frac{\pi}{2}) = 2$, and $A(1.19606) \approx 2.2018$, which is the absolute maximum possible cross-sectional area.

3.3.2 Summary

- 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.

3.3.3 Exercises

1. (Maximizing the volume of a box) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

2. (Minimizing the cost of a container) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Maximizing area contained by a fence) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Minimizing the area of a poster) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Maximizing the area of a rectangle) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. 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 top and 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.

Answer. The absolute maximum volume is $V\left(\sqrt{\frac{5}{3}}\right) = \frac{15}{12}\left(\sqrt{\frac{5}{3}}\right) - \frac{1}{4}\left(\sqrt{\frac{5}{3}}\right)^3 \approx 1.07583$ cubic feet.

Solution. Let x be the length of one side of the square bottom and h the height of one of the sides. We first note that the volume, V , of the box is $V = x^2h$, and the surface area of the box is $S = 2x^2 + 4xh$, since there are two square sides (bottom and top) of area x^2 and four identical sides of area xh . Since the material for the top and bottom costs \$3.00 per square foot and the material for the sides costs \$1.50 per square foot, it follows that the total cost of a box with surface area S is

$$C = 1.50(2x^2) + 3.00(4xh) = 3x^2 + 12xh.$$

We are interested in building box of largest volume for which we spend \$15. This cost constraint implies that

$$15 = 3x^2 + 12xh,$$

which allows us to express h as a function of x by solving for h . Doing so, we see that $12xh = 15 - 3x^2$, so

$$h = \frac{15 - 3x^2}{12x}.$$

Using this relationship, we are now able to express the volume, V , of the box as a function of the single variable x . Substituting the expression for h , we have

$$V = x^2h = x^2 \frac{15 - 3x^2}{12x} = \frac{1}{12}x(15 - 3x^2) = \frac{15}{12}x - \frac{1}{4}x^3.$$

Note particularly that in order to have a box, we require that both $x > 0$ and $h > 0$. The latter condition along with the fact that $h = \frac{15-3x^2}{12x}$ shows that we need $15 - 3x^2 > 0$, and thus $x^2 < 5$, so $x < \sqrt{5}$. Hence we are considering $V(x)$ on the domain $0 < x < \sqrt{5}$.

It is apparent that the box vanishes (and its volume tends to zero) as either $x \rightarrow 0^+$ or $x \rightarrow \sqrt{5}^-$, and thus the absolute maximum volume must appear between these values at a critical number. Differentiating V , we find that $V'(x) = \frac{15}{12} - \frac{3}{4}x^2$. Solving $V'(x) = 0$ to find all critical numbers of V in the domain, we see that $x^2 = \frac{15}{12} \cdot \frac{4}{3} = \frac{5}{3}$, and thus $x = \pm\sqrt{\frac{5}{3}}$, so only $x = \sqrt{\frac{5}{3}}$ is a critical number that lies in $0 < x < \sqrt{5}$. Noting that the absolute minimum volume

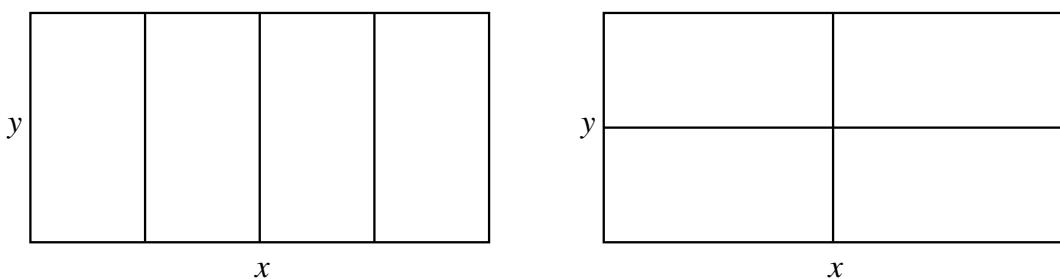
$V = 0$ occurs at each endpoint, it must be the case that the absolute maximum volume is

$$V\left(\sqrt{\frac{5}{3}}\right) = \frac{15}{12}\left(\sqrt{\frac{5}{3}}\right) - \frac{1}{4}\left(\sqrt{\frac{5}{3}}\right)^3 \approx 1.07583 \text{ cubic feet.}$$

7. 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 by adding three segments of fence inside the large rectangle that are parallel to two existing sides. She has decided to purchase 7500 ft of fencing. What is the maximum possible area that each of the four pens will enclose?

Answer. Exercise Answer

Solution. The farmer will fence a rectangular area whose length x and width y are unknown. The total area of the rectangular area is composed of 4 equal area rectangles. There are two ways to do this as shown in the figure below.



We want to maximize the area of each subrectangular region, which is the same as maximizing the total area $A = xy$ of the entire rectangular region. Since A has two variables, we can't yet directly apply calculus techniques to it. We use the information about the amount of fencing to relate x and y in order to write A in terms of just one variable.

In the case of the fencing option on the left in the figure, the total amount of fencing that is needed to enclose the four grazing areas is $2x + 5y$. We assume that all of the 7500 feet of fencing will be used, so $2x + 5y = 7500$. Solving for x , $x = 3750 - 2.5y$. Substituting this expression for x into the area formula allows us to write A in terms of y alone:

$$A(y) = (3750 - 2.5y)y = 3750y - 2.5y^2.$$

To enclose any area, we of course must have $y > 0$ and $x > 0$, and since $x = 3750 - 2.5y$, it follows that we need to have $2.5y < 3750$ so $y < 1500$. To find the maximum value of A on this domain, we find the critical numbers of A . Since $A'(y) = 3750 - 5y$, we see that $A'(y) = 0$ when and only when $y = 750$. For $0 < y < 750$, we see that $A'(y) > 0$ and for $750 < y < 1500$ we have $A'(y) < 0$. So $A(750) = 1406250$ square feet is the absolute maximum value of A . The maximum possible area that each of the four pens can enclose is one quarter of the total, or 351562.5 square feet.

In the case of the fencing option on the right in the figure, the total amount of fencing that is needed to enclose the four grazing areas is $3x + 3y$. We assume that all of the 7500 feet of fencing will be used, so $3x + 3y = 7500$. Solving for x and substituting into A , we find $x = 2500 - y$ and

$$A(y) = (2500 - y)y = 2500y - y^2.$$

To enclose any area, we of course must have $y > 0$ and $x > 0$, and since $x = 2500 - y$, it follows that we need to have $y < 2500$. To find the maximum value of A on this domain,

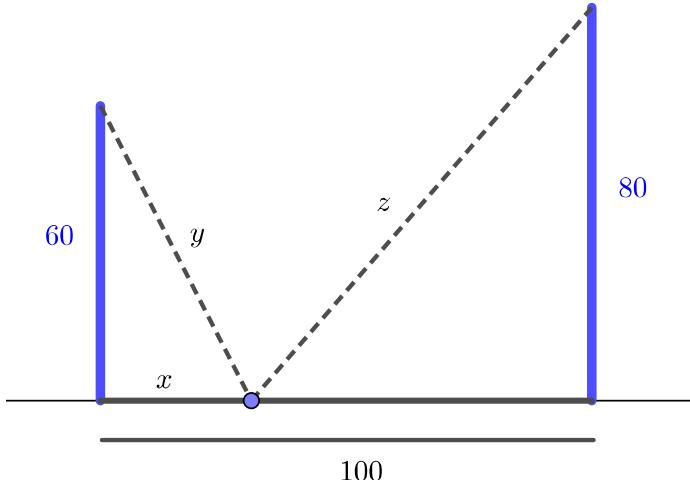
we find the critical numbers of A . Since $A'(y) = 2500 - 2y$, we see that $A'(y) = 0$ when and only when $y = 1250$. For $0 < y < 1250$, we see that $A'(y) > 0$ and for $1250 < y < 2500$ we have $A'(y) < 0$. So $A(1250) = 1406250$ square feet is the absolute maximum value of A . The maximum possible area that each of the four pens can enclose is one quarter of the total, or 390625 square feet.

So the maximum possible area that each of the four pens can enclose is 390625 square feet.

8. Two vertical poles 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.

Answer. 172.047 feet of cable.

Solution. We start by constructing a careful diagram of the overall situation. Let x be the distance from the first pole to the second pole where the cable touches the ground, and let y be the length of the cable from the top of the first tower to the ground, and z the length of cable from the ground to the top of the second tower, as pictured in the following figure. Note that we have included the given heights of the towers and the distance between them.



The total amount of cable, L , used (which we wish to minimize) is given by $L = y + z$. We can use the geometry of the situation to write both y and z in terms of x . In the first right triangle, by the Pythagorean Theorem we know that $x^2 + 60^2 = y^2$, and thus $y = \sqrt{x^2 + 3600}$. In the second right triangle, note that the horizontal leg of the triangle has length $100 - x$, and thus $(100 - x)^2 + 80^2 = z^2$, so $z = \sqrt{(100 - x)^2 + 6400}$. It follows that with $L = y + z$,

$$L(x) = \sqrt{x^2 + 3600} + \sqrt{6400 + (100 - x)^2}.$$

It is also apparent from the given situation that the relevant domain for L is $0 \leq x \leq 100$. Given that we want to minimize L on this closed interval, it remains to find any critical

numbers of L .

We note by the chain rule that

$$L'(x) = \frac{2x}{2(x^2 + 3600)^{1/2}} + \frac{-2(100 - x)}{2((100 - x)^2 + 6400)^{1/2}}$$

We use technology to solve the equation $L'(x) = 0$ and find that it has a single solution in the interval $0 \leq x \leq 100$: $x \approx 42.85714$.

To find the absolute minimum of L on the relevant domain, we evaluate L at the endpoints and the only critical number. Doing so, we find $L(0) \approx 188.062$, $L(100) \approx 196.619$, and $L(42.85714) \approx 172.047$. Hence the least amount of cable we can use is approximately 172.047 feet and is accomplished by anchoring the cable about 42.85714 feet from the first tower.

9. 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.

Answer. The minimum cost is \$1165.70.

Solution. We let the length of the cylindrical part of the tank be h with the radius of the cylinder given by r . It follows that the radius of the hemispherical ends of the tank is also r . We first identify formulas for both the volume and surface area of the tank by looking at the cylindrical and hemispherical pieces.

The cylindrical part of the tank has volume $V_c = \pi r^2 h$, and the surface area of its sides is $A_c = 2\pi r h$ (which is the circumference of the cylinder's base times the cylinder's height – think about how a cylinder unfurls to form a rectangle). The two hemispherical ends together constitute a full sphere, and the volume of a sphere is $V_s = \frac{4}{3}\pi r^3$, while the surface area of a sphere is $A_s = 4\pi r^2$.

The overall volume of the tank is therefore

$$V = V_c + V_s = \pi r^2 h + \frac{4}{3}\pi r^3,$$

and its total surface area is

$$A = A_c + A_s = 2\pi r h + 4\pi r^2.$$

Moreover, taking into account the different costs of material for the cylindrical sides and hemispherical ends, we find that the total cost of the tank is

$$C = 2(2\pi r h) + 5(4\pi r^2) = 4\pi r h + 20\pi r^2.$$

Our goal is to minimize the total cost of the tank given that we want it to hold 1000 cubic meters of volume. From this constraint, we have the equation $1000 = \pi r^2 h + \frac{4}{3}\pi r^3$, which we can solve for h to get $\pi r^2 h = 1000 - \frac{4}{3}\pi r^3$, so

$$h = \frac{1000 - \frac{4}{3}\pi r^3}{\pi r^2}.$$

This allows us to now write cost as a function of the single variable r by substituting the expression for h into the earlier formula we found for cost. In particular,

$$C(r) = 4\pi r \frac{1000 - \frac{4}{3}\pi r^3}{\pi r^2} + 20\pi r^2 = 4 \frac{1000 - \frac{4}{3}\pi r^3}{r} + 20\pi r^2$$

Simplifying further, we have found that

$$C(r) = \frac{4000}{r} - \frac{16}{3}\pi r^2 + 20\pi r^2 = \frac{4000}{r} + \frac{44}{3}r^2.$$

Based on the physical constraints of the problem, we obviously will only consider values of $r > 0$ and $h > 0$. Since $h = \frac{1000 - \frac{4}{3}\pi r^3}{\pi r^2}$, we must have

$$\frac{4}{3}\pi r^3 < 1000$$

so $r^3 < \frac{3}{4} \cdot 1000 = 750$, and thus $r < \sqrt[3]{750} \approx 9.086$. We now work to minimize $C(r)$ on the interval $0 < r < \sqrt[3]{750}$.

Taking the derivative of C , we find that $C'(r) = -\frac{4000}{r^2} + \frac{88}{3}r$. Setting $C'(r) = 0$, it follows that

$$\frac{4000}{r^2} = \frac{88}{3}r$$

and thus

$$4000 \cdot \frac{3}{88} = r^3.$$

Therefore, $r^3 = \frac{1500}{11}$, so $r = \sqrt[3]{\frac{1500}{11}} \approx 5.147$ is the only critical number of C ; observe that it lies within the interval $0 < r < \sqrt[3]{750}$.

If we plot the derivative function, $y = C'(r)$, on the interval $0 < r < \sqrt[3]{750}$, we see that for $0 < r < 5.147$, $C'(r) < 0$, and for $5.147 < r < 9.086$, $C'(r) > 0$, which shows that C is decreasing before the critical number and increasing after the critical number, and hence C has a global minimum of $C(\sqrt[3]{\frac{1500}{11}}) \approx 1165.70$, so \$1165.70 is the minimum cost to build the desired tank under the given constraints.

3.4 Using derivatives to describe families of functions

Motivating Questions

- Given a family of functions that depends on one or more parameters, how does the shape of the graph of a typical function in the family depend on the value of the parameters?
 - How can we construct first and second derivative sign charts of functions that depend on one or more parameters while allowing those parameters to remain arbitrary constants?
-

Mathematicians are often interested in making general observations, say by describing patterns that hold in a large number of cases. Think about the Pythagorean Theorem: it doesn't tell us something about a single right triangle, but rather a fact about *every* right triangle. In the next part of our studies, we use calculus to make general observations about families of functions that depend on one or more parameters. People who use applied mathematics, such as engineers and economists, often encounter the same types of functions where only small changes to certain constants occur. These constants are called *parameters*.

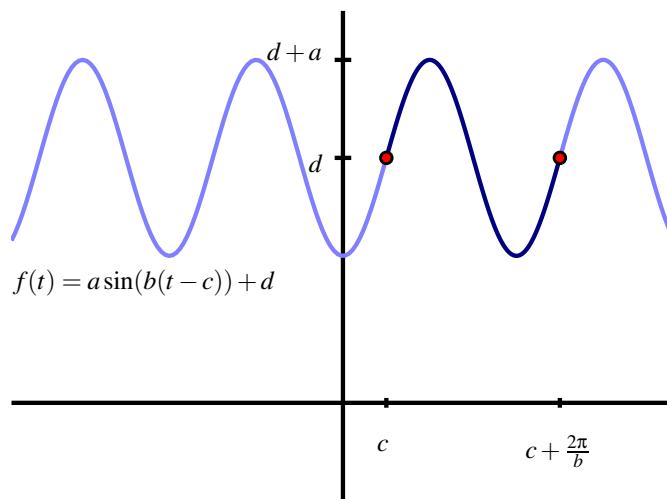


Figure 3.4.1: The graph of $f(t) = a \sin(b(t - c)) + d$ based on parameters a , b , c , and d .

You are already familiar with certain families of functions. For example, $f(t) = a \sin(b(t - c)) + d$ is a stretched and shifted version of the sine function with amplitude a , period $\frac{2\pi}{b}$, phase shift c , and vertical shift d . We know that a affects the size of the oscillation, b the rapidity of oscillation, and c where the oscillation starts, as shown in Figure 3.4.1, while d

affects the vertical positioning of the graph.

As another example, every function of the form $y = mx + b$ is a line with slope m and y -intercept $(0, b)$. The value of m affects the line's steepness, and the value of b situates the line vertically on the coordinate axes. These two parameters describe all possible non-vertical lines.

For other less familiar families of functions, we can use calculus to discover where key behavior occurs: where members of the family are increasing or decreasing, concave up or concave down, where relative extremes occur, and more, all in terms of the parameters involved. To get started, we revisit a common collection of functions to see how calculus confirms things we already know.

Example 3.4.2. Let a, h , and k be arbitrary real numbers with $a \neq 0$, and let f be the function given by the rule $f(x) = a(x - h)^2 + k$.

- a. What familiar type of function is f ? What information do you know about f just by looking at its form? (Think about the roles of a, h , and k .)
- b. Next we use some calculus to develop familiar ideas from a different perspective. To start, treat a, h , and k as constants and compute $f'(x)$.
- c. Find all critical numbers of f . (These will depend on at least one of a, h , and k .)
- d. Assume that $a < 0$. Construct a first derivative sign chart for f .
- e. Based on the information you've found above, classify the critical values of f as maxima or minima.

3.4.1 Describing families of functions in terms of parameters

Our goal is to describe the key characteristics of the overall behavior of each member of a family of functions in terms of its parameters. By finding the first and second derivatives and constructing sign charts (each of which may depend on one or more of the parameters), we can often make broad conclusions about how each member of the family will appear.

Example 3.4.3. Consider the two-parameter family of functions given by $g(x) = axe^{-bx}$, where a and b are positive real numbers. Fully describe the behavior of a typical member of the family in terms of a and b , including the location of all critical numbers, where g is increasing, decreasing, concave up, and concave down, and the long term behavior of g .

Solution. We begin by computing $g'(x)$. By the product rule,

$$g'(x) = ax \frac{d}{dx} [e^{-bx}] + e^{-bx} \frac{d}{dx} [ax].$$

By applying the chain rule and constant multiple rule, we find that

$$g'(x) = axe^{-bx}(-b) + e^{-bx}(a).$$

To find the critical numbers of g , we solve the equation $g'(x) = 0$. By factoring $g'(x)$, we find

$$0 = ae^{-bx}(-bx + 1).$$

Since we are given that $a \neq 0$ and we know that $e^{-bx} \neq 0$ for all values of x , the only way this equation can hold is when $-bx + 1 = 0$. Solving for x , we find $x = \frac{1}{b}$, and this is therefore the only critical number of g .

We construct the first derivative sign chart for g that is shown in Figure 3.4.4.

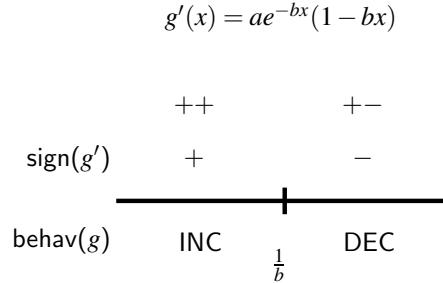


Figure 3.4.4: The first derivative sign chart for $g(x) = axe^{-bx}$.

Because the factor ae^{-bx} is always positive, the sign of g' depends on the linear factor $(1 - bx)$, which is positive for $x < \frac{1}{b}$ and negative for $x > \frac{1}{b}$. Hence we can not only conclude that g is always increasing for $x < \frac{1}{b}$ and decreasing for $x > \frac{1}{b}$, but also that g has a global maximum at $(\frac{1}{b}, g(\frac{1}{b}))$ and no local minimum.

We turn next to analyzing the concavity of g . With $g'(x) = -abxe^{-bx} + ae^{-bx}$, we differentiate to find that

$$g''(x) = -abxe^{-bx}(-b) + e^{-bx}(-ab) + ae^{-bx}(-b).$$

Combining like terms and factoring, we now have

$$g''(x) = ab^2xe^{-bx} - 2abe^{-bx} = abe^{-bx}(bx - 2).$$

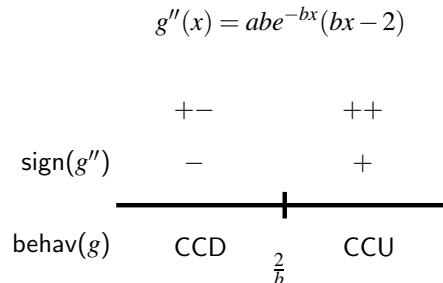


Figure 3.4.5: The second derivative sign chart for $g(x) = axe^{-bx}$.

We observe that abe^{-bx} is always positive, and thus the sign of g'' depends on the sign of

$(bx - 2)$, which is zero when $x = \frac{2}{b}$. Since b is positive, the value of $(bx - 2)$ is negative for $x < \frac{2}{b}$ and positive for $x > \frac{2}{b}$. The sign chart for g'' is shown in Figure 3.4.5. Thus, g is concave down for all $x < \frac{2}{b}$ and concave up for all $x > \frac{2}{b}$.

Finally, we analyze the long term behavior of g by considering two limits. First, we note that

$$\lim_{x \rightarrow \infty} g(x) = \lim_{x \rightarrow \infty} axe^{-bx} = \lim_{x \rightarrow \infty} \frac{ax}{e^{bx}}.$$

This limit has indeterminate form $\frac{\infty}{\infty}$, so we apply L'Hopital's Rule and find that $\lim_{x \rightarrow \infty} g(x) = 0$. In the other direction,

$$\lim_{x \rightarrow -\infty} g(x) = \lim_{x \rightarrow -\infty} axe^{-bx} = -\infty,$$

because $ax \rightarrow -\infty$ and $e^{-bx} \rightarrow \infty$ as $x \rightarrow -\infty$. Hence, as we move left on its graph, g decreases without bound, while as we move to the right, $g(x) \rightarrow 0$.

All of this information now helps us produce the graph of a typical member of this family of functions without using a graphing utility (and without choosing particular values for a and b), as shown in Figure 3.4.6.

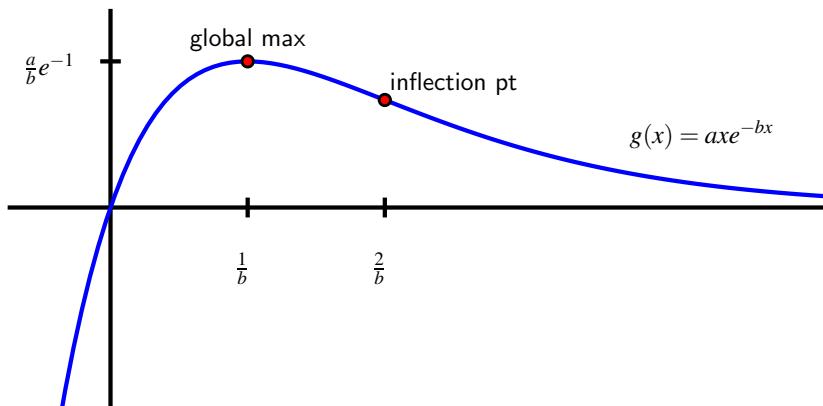


Figure 3.4.6: The graph of $g(x) = axe^{-bx}$.

Note that the value of b controls the horizontal location of the global maximum and the inflection point, as neither depends on a . The value of a affects the vertical stretch of the graph. For example, the global maximum occurs at the point $(\frac{1}{b}, g(\frac{1}{b})) = (\frac{1}{b}, \frac{a}{b}e^{-1})$, so the larger the value of a , the greater the value of the global maximum.

The work we've completed in Example 3.4.3 can often be replicated for other families of functions that depend on parameters. Normally we are most interested in determining all critical numbers, a first derivative sign chart, a second derivative sign chart, and the limit of the function as $x \rightarrow \infty$. Throughout, we prefer to work with the parameters as arbitrary constants. In addition, we can experiment with some particular values of the parameters present to reduce the algebraic complexity of our work. The following activities offer several key examples where we see that the values of the parameters substantially affect the behavior of individual functions within a given family.

Example 3.4.7. Consider the family of functions defined by $p(x) = x^3 - ax$, where $a \neq 0$ is an arbitrary constant.

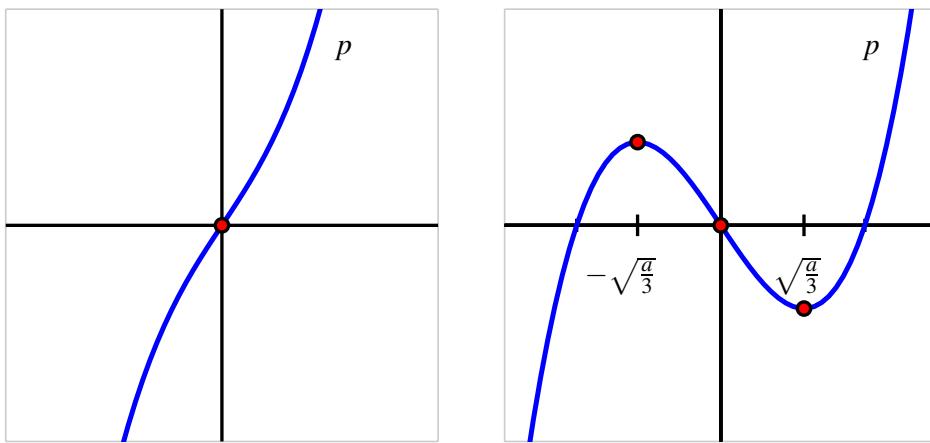
- a. Find $p'(x)$ and determine the critical numbers of p . How many critical numbers does p have?
- b. Construct a first derivative sign chart for p . What can you say about the overall behavior of p if the constant a is positive? Why? What if the constant a is negative? In each case, describe the relative extremes of p .
- c. Find $p''(x)$ and construct a second derivative sign chart for p . What does this tell you about the concavity of p ? What role does a play in determining the concavity of p ?
- d. Without using a graphing utility, sketch and label typical graphs of $p(x)$ for the cases where $a > 0$ and $a < 0$. Label all inflection points and local extrema.
- e. Finally, use a graphing utility to test your observations above by entering and plotting the function $p(x) = x^3 - ax$ for at least four different values of a . Write several sentences to describe your overall conclusions about how the behavior of p depends on a .

Hint.

- a. When solving $p'(x) = 0$, think about two possible cases: when $a > 0$ and when $a < 0$.
- b. Remember that any quadratic function can be zero at most two times. How does the graph of $y = 3x^2 - a$ look?
- c. Don't forget that $\frac{d}{dx}[a] = 0$.
- d. Think about how a typical cubic polynomial's graph behaves.

Answer.

- a. p has two critical numbers ($x = \pm\sqrt{\frac{a}{3}}$) whenever $a > 0$ and no critical numbers when $a < 0$.
- b. When $a < 0$, p is always increasing and has no relative extreme values. When $a > 0$, p has a relative maximum at $x = -\sqrt{\frac{a}{3}}$ and a relative minimum at $x = +\sqrt{\frac{a}{3}}$.
- c. p is CCD for $x < 0$ and p is CCU for $x > 0$, making $x = 0$ an inflection point.
- d.

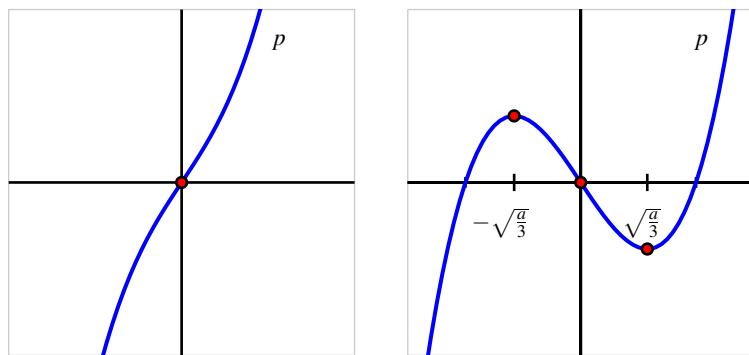
**Solution.**

- a. We first note that $p'(x) = 3x^2 - a$, so to find critical numbers we set $p'(x) = 0$ and solve for x . This leads to the equation $3x^2 - a = 0$, which implies

$$x^2 = \frac{a}{3}.$$

If $a > 0$, then the solutions to this equation are $x = \pm\sqrt{\frac{a}{3}}$; if $a < 0$, then the equation has no solution. Hence, p has two critical numbers ($x = \pm\sqrt{\frac{a}{3}}$) whenever $a > 0$ and no critical numbers when $a < 0$.

- b. For the case when $a < 0$, we observe that $p'(x) = 3x^2 - a$ is positive for every value of x , and thus p is always increasing and has no relative extreme values. (There are no critical numbers to place on the first derivative sign chart, and p' is always positive.) For the case when $a > 0$, we observe that $p'(x) = 3x^2 - a$ is a concave up parabola with zeros at $x = -\sqrt{\frac{a}{3}}$ and $x = +\sqrt{\frac{a}{3}}$. It follows that for $x < -\sqrt{\frac{a}{3}}$, $p'(x) > 0$ (so p is increasing); for $-\sqrt{\frac{a}{3}} < x < \sqrt{\frac{a}{3}}$, $p'(x) < 0$ (so p is decreasing); and for $x > \sqrt{\frac{a}{3}}$, $p'(x) > 0$ (so p is again increasing). In this situation, we see that p has a relative maximum at $x = -\sqrt{\frac{a}{3}}$ and a relative minimum at $x = +\sqrt{\frac{a}{3}}$.
- c. Since $p'(x) = 3x^2 - a$ and a is constant, it follows that $p''(x) = 6x$. Note that $p''(x) = 0$ when $x = 0$ and that $p''(x) < 0$ for $x < 0$ and $p''(x) > 0$ for $x > 0$. Hence p is CCD for $x < 0$ and p is CCU for $x > 0$, making $x = 0$ an inflection point.
- d. Below, we show the two possible situations. At left, for the case when $a < 0$ and p is always increasing with an inflection point at $x = 0$, and at right for when $a > 0$ and p has a relative maximum at $x = -\sqrt{\frac{a}{3}}$ and a relative minimum at $x = +\sqrt{\frac{a}{3}}$, again with an inflection point at $x = 0$. Note, too, that p has its x -intercepts at $x = \pm\sqrt{a}$.



Example 3.4.8. Consider the two-parameter family of functions of the form $h(x) = a(1 - e^{-bx})$, where a and b are positive real numbers.

- Find the first derivative and the critical numbers of h . Use these to construct a first derivative sign chart and determine for which values of x the function h is increasing and decreasing.
- Find the second derivative and build a second derivative sign chart. For which values of x is a function in this family concave up? concave down?
- What is the value of $\lim_{x \rightarrow \infty} a(1 - e^{-bx})$? $\lim_{x \rightarrow -\infty} a(1 - e^{-bx})$?
- How does changing the value of b affect the shape of the curve?
- Without using a graphing utility, sketch the graph of a typical member of this family. Write several sentences to describe the overall behavior of a typical function h and how this behavior depends on a and b .

Hint.

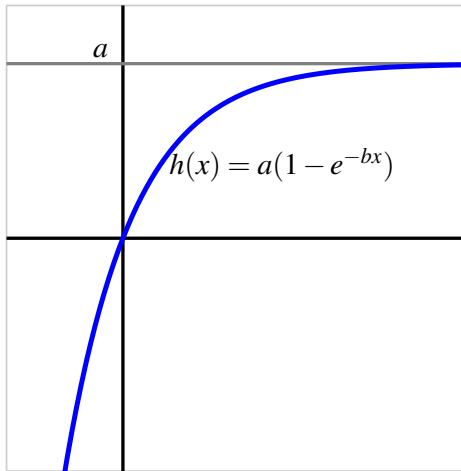
- Expand to write $h(x) = a - ae^{-bx}$ before differentiating.
- Remember that e^{-bx} is never zero and always positive, regardless of the value of x .
- Recall that $e^{-x} \rightarrow 0$ as $x \rightarrow \infty$ and $e^{-x} \rightarrow \infty$ as $x \rightarrow -\infty$.
- Consider how b affects the value of $h'(x)$.
- Use your work in (a)-(d).

Answer.

- h is an always increasing function.
- h is always concave down.
- $\lim_{x \rightarrow \infty} a(1 - e^{-bx}) = a$, and $\lim_{x \rightarrow -\infty} a(1 - e^{-bx}) = -\infty$.

- d. If b is large and x is close to zero, $h'(x)$ is relatively large near $x = 0$, and the curve's slope will quickly approach zero as x increases. If b is small, the graph is less steep near $x = 0$ and its slope goes to zero less quickly as x increases.

e.



Solution.

- a. Since $h(x) = a - ae^{-bx}$, we have by the constant multiple and chain rules that

$$h'(x) = -ae^{-bx}(-b) = abe^{-bx}.$$

Since a and b are positive constants and $e^{-bx} > 0$ for all x , we see that $h'(x)$ is never zero (nor undefined), and indeed $h'(x) > 0$ for all x . Hence h is an always increasing function.

- b. Because $h'(x) = abe^{-bx}$, we have that $h''(x) = abe^{-bx}(-b) = -ab^2e^{-bx}$. As with h' , we recognize that a , b^2 , and e^{-bx} are always positive, and thus $h''(x) = -ab^2e^{-bx} < 0$ for all values of x , making h always concave down.

- c. As $x \rightarrow \infty$, $e^{-bx} \rightarrow 0$. Thus,

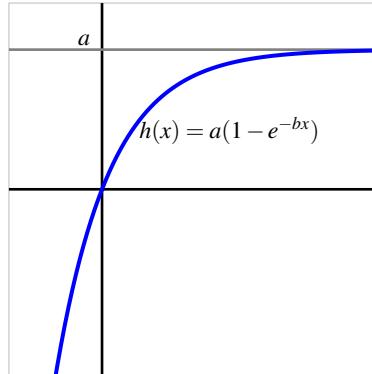
$$\lim_{x \rightarrow \infty} a(1 - e^{-bx}) = \lim_{x \rightarrow \infty} a - ae^{-bx} = a - 0 = a.$$

This shows that h has a horizontal asymptote at $y = a$ as we move rightward on its graph. As $x \rightarrow -\infty$, $e^{-bx} \rightarrow \infty$. Thus,

$$\lim_{x \rightarrow -\infty} a(1 - e^{-bx}) = \lim_{x \rightarrow -\infty} a - ae^{-bx} = -\infty.$$

- d. Noting that $h'(x) = abe^{-bx}$, we see that if we consider different values of b , the slope of the graph changes. If b is large and x is close to zero, $h'(x) \approx ab$ (since $e^0 = 1$), so $h'(x)$ is relatively large near $x = 0$. At the same time, for large b , e^{-bx} approaches zero quickly as x increases, so the curve's slope will quickly approach zero as x increases. If b is small, the graph is less steep near $x = 0$ and its slope goes to zero less quickly as x increases.

- e. Observing that $h(0) = 0$ and $\lim_{x \rightarrow \infty} h(x) = a$, along with the facts that h is always increasing and always concave down, we see that a typical member of this family looks like the following graph.



Example 3.4.9. Let $L(t) = \frac{A}{1+ce^{-kt}}$, where A , c , and k are all positive real numbers.

- Observe that we can equivalently write $L(t) = A(1 + ce^{-kt})^{-1}$. Find $L'(t)$ and explain why L has no critical numbers. Is L always increasing or always decreasing? Why?
- Given the fact that
$$L''(t) = Ack^2e^{-kt} \frac{ce^{-kt} - 1}{(1 + ce^{-kt})^3},$$
find all values of t such that $L''(t) = 0$ and hence construct a second derivative sign chart. For which values of t is a function in this family concave up? concave down?
- What is the value of $\lim_{t \rightarrow \infty} \frac{A}{1+ce^{-kt}}$? $\lim_{t \rightarrow -\infty} \frac{A}{1+ce^{-kt}}$?
- Find the value of $L(x)$ at the inflection point found in (b).
- Without using a graphing utility, sketch the graph of a typical member of this family. Write several sentences to describe the overall behavior of a typical function L and how this behavior depends on A , c , and k number.
- Explain why it is reasonable to think that the function $L(t)$ models the growth of a population over time in a setting where the largest possible population the surrounding environment can support is A .

Hint.

- Use the chain rule, treating A , c , and k as constants.
- Note that the only way $L''(t) = 0$ is if $ce^{-kt} - 1 = 0$.
- Remember that $e^{-t} \rightarrow 0$ as $t \rightarrow \infty$ and $e^{-t} \rightarrow \infty$ as $t \rightarrow -\infty$.
- Don't forget that $e^{\ln(x)} = x$ for all $x > 0$.

- e. Think about horizontal asymptotes, where L is increasing and decreasing, and concavity.

Answer.

- a. L is an always increasing function.

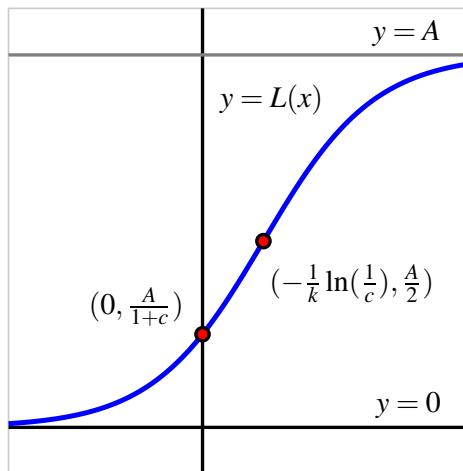
- b. L is concave up for all $t < -\frac{1}{k} \ln(\frac{1}{c})$ and concave down for all other values of t .

- c. $\lim_{t \rightarrow \infty} \frac{A}{1+ce^{-kt}} = A$, and

$$\lim_{t \rightarrow \infty} \frac{A}{1+ce^{-kt}} = 0.$$

- d. The inflection point on the graph of L is $(-\frac{1}{k} \ln(\frac{1}{c}), \frac{A}{2})$.

- e.



Solution.

- a. By the chain rule and treating A , c , and k as constants, we find that

$$L'(t) = A(-1)(1+ce^{-kt})^{-2}ce^{-kt}(-k) = Acke^{-kt}(1+ce^{-kt})^{-2}.$$

Since A , c , and k are all positive and $e^{-kt} > 0$ for all values of t , it is apparent that $L'(t)$ is never zero, and indeed is positive for every value of t . Thus, L is an always increasing function.

- b. Given that

$$L''(t) = Ack^2e^{-kt} \frac{ce^{-kt} - 1}{(1+ce^{-kt})^3},$$

the only way $L''(t) = 0$ is if $ce^{-kt} - 1 = 0$. Solving $ce^{-kt} - 1 = 0$ for t , we first write $e^{-kt} = \frac{1}{c}$. Taking the natural logarithm of both sides, $-kt = \ln(\frac{1}{c})$, so that

$$t = -\frac{1}{k} \ln\left(\frac{1}{c}\right)$$

is the only value of t for which $L''(t) = 0$. Now, observe that since $ce^{-kt} \rightarrow 0$ as $t \rightarrow \infty$, the quantity $ce^{-kt} - 1$ will be positive to the left of where it is zero and negative to the right of where it is zero. Since this is the only term in $L''(t)$ that can change sign, it follows that $L''(t) > 0$ for $t < -\frac{1}{k} \ln(\frac{1}{c})$ and $L''(t) > 0$ for $t > -\frac{1}{k} \ln(\frac{1}{c})$, making L concave up to the left of the noted inflection point and concave down thereafter.

- c. Recalling that $e^{-kt} \rightarrow 0$ as $t \rightarrow \infty$, we observe that

$$\lim_{t \rightarrow \infty} \frac{A}{1 + ce^{-kt}} = \frac{A}{1 + 0} = A,$$

so L has a horizontal asymptote of $y = A$ as $t \rightarrow \infty$. On the other hand, since $e^{-kt} \rightarrow \infty$ as $t \rightarrow -\infty$, this causes the denominator of L to grow without bound (while the numerator remains constant), and therefore

$$\lim_{t \rightarrow -\infty} \frac{A}{1 + ce^{-kt}} = 0,$$

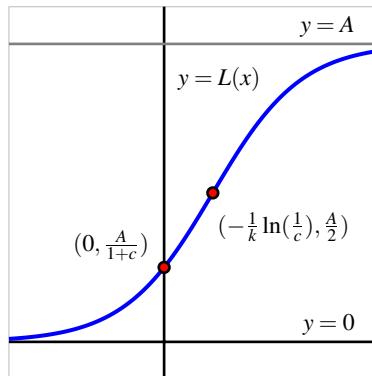
which means L has a horizontal asymptote of $y = 0$ as $t \rightarrow -\infty$.

- d. From (b), we know that $t = -\frac{1}{k} \ln(\frac{1}{c})$ is the location of the inflection point of L . Evaluating the function at this point, we find that

$$L\left(-\frac{1}{k} \ln\left(\frac{1}{c}\right)\right) = \frac{A}{1 + ce^{-k[-\frac{1}{k} \ln(\frac{1}{c})]}} = \frac{A}{1 + ce^{\ln(\frac{1}{c})}} = \frac{A}{1 + c \cdot \frac{1}{c}} = \frac{A}{2}.$$

Thus, the inflection point on the graph of L is located at $(-\frac{1}{k} \ln(\frac{1}{c}), \frac{A}{2})$.

- e. We have shown that L is an always increasing function that has horizontal asymptotes at $y = 0$ and $y = A$, as well as an inflection point at $(-\frac{1}{k} \ln(\frac{1}{c}), \frac{A}{2})$, which we note lies vertically halfway between the asymptotes. In addition, we see that $L(0) = \frac{A}{1+c}$. The combination of all of this information shows us that a typical graph in this family of functions is given by the following figure.



3.4.2 Summary

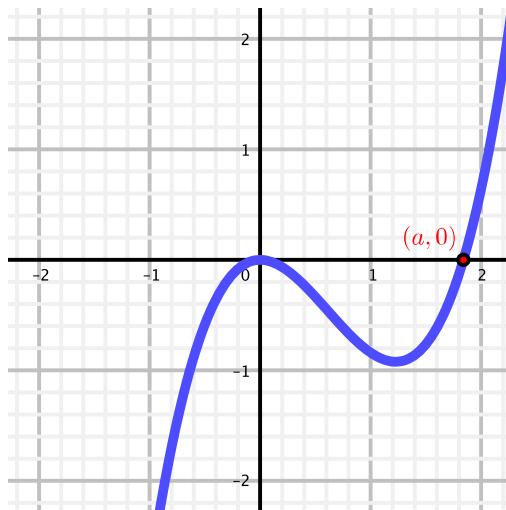
- Given a family of functions that depends on one or more parameters, by investigating how critical numbers and locations where the second derivative is zero depend on the values of these parameters, we can often accurately describe the shape of the function in terms of the parameters.
- In particular, just as we can create first and second derivative sign charts for a single function, we often can do so for entire families of functions where critical numbers and possible inflection points depend on arbitrary constants. These sign charts then reveal where members of the family are increasing or decreasing, concave up or concave down, and help us to identify relative extremes and inflection points.

3.4.3 Exercises

- (Drug dosage with a parameter) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Using the graph of g') This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- Consider the one-parameter family of functions given by $p(x) = x^3 - ax^2$, where $a > 0$.
 - Sketch a plot of a typical member of the family, using the fact that each is a cubic polynomial with a repeated zero at $x = 0$ and another zero at $x = a$.
 - Find all critical numbers of p .
 - Compute p'' and find all values for which $p''(x) = 0$. Hence construct a second derivative sign chart for p .
 - Describe how the location of the critical numbers and the inflection point of p change as a changes. That is, if the value of a is increased, what happens to the critical numbers and inflection point?

Answer.

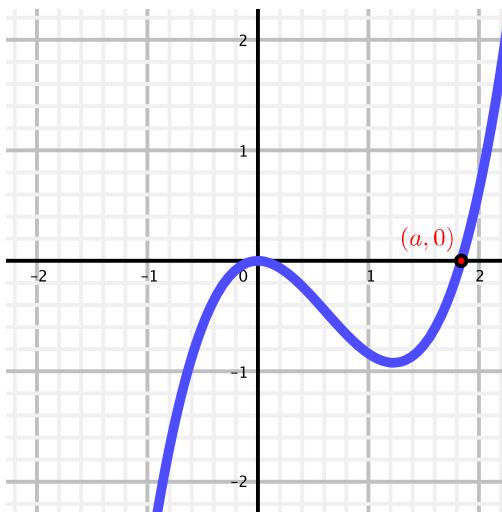
a.



- b. $x = 0$ and $x = \frac{2a}{3}$.
- c. $x = \frac{a}{2}$; $p''(x)$ changes sign from negative to positive at $x = \frac{a}{2}$.
- d. As we increase the value of a , both the location of the critical number and the inflection point move to the right along with a .

Solution.

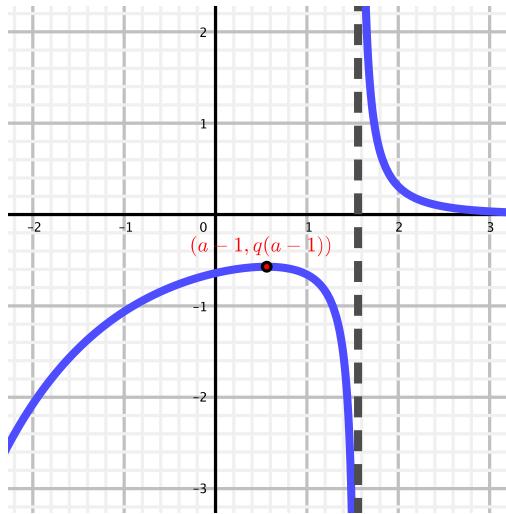
- a. Writing $p(x) = x^3 - ax^2 = x^2(x - a)$, we see that $p(x) = 0$ at $x = 0$ and $x = a$. Further, since $x^2 \geq 0$ for all x , we note that $p(x) \leq 0$ whenever $x < a$ and $p(x) > 0$ for all $x > a$. A typical such function is plotted below.



- b. To find the critical numbers of p , we first compute $p'(x)$. Treating a as constant, we have $p'(x) = 3x^2 - 2ax = x(3x - 2a)$. Thus, setting $p'(x) = 0$, we find that $x(3x - 2a) = 0$ and thus $x = 0$ and $x = \frac{2a}{3}$ are the critical numbers of p .
- c. Observe that $p''(x) = 6x - 2a$ and thus $p''(x) = 0$ implies that $x = \frac{a}{3}$. Since $p''(x)$ is linear with positive slope, we see that $p''(x)$ changes sign from negative to positive at $x = \frac{a}{3}$, and thus p is concave down for $x < \frac{a}{3}$ and concave up for $x > \frac{a}{3}$ and $x = \frac{a}{3}$ is the only inflection point of p .
- d. There is always a critical number of p at $x = 0$. But the other critical number at $x = \frac{2a}{3}$ and the inflection point at $x = \frac{a}{3}$ both obviously depend directly on a : as we increase the value of a , both the location of the critical number and the inflection point move to the right along with a .
- 4. Let $q(x) = \frac{e^{-x}}{x-c}$ be a one-parameter family of functions where $c > 0$.
 - a. Explain why q has a vertical asymptote at $x = c$.
 - b. Determine $\lim_{x \rightarrow \infty} q(x)$ and $\lim_{x \rightarrow -\infty} q(x)$.
 - c. Compute $q'(x)$ and find all critical numbers of q .
 - d. Construct a first derivative sign chart for q and determine whether each critical number leads to a local minimum, local maximum, or neither for the function q .
 - e. Sketch a typical member of this family of functions with important behaviors clearly labeled.

Answer.

- a. $x = c$ is a vertical asymptote because $\lim_{x \rightarrow c^+} \frac{e^{-x}}{x-c} = \infty$ and $\lim_{x \rightarrow c^-} \frac{e^{-x}}{x-c} = -\infty$.
- b. $\lim_{x \rightarrow \infty} \frac{e^{-x}}{x-c} = 0$; $\lim_{x \rightarrow -\infty} \frac{e^{-x}}{x-c} = -\infty$.
- c. The only critical number for q is $x = c - 1$.
- d. When $x < c - 1$, $q'(x) > 0$; when $x > c - 1$, $q'(x) < 0$; q has a local maximum at $x = c - 1$.
- e.

**Solution.**

- a. The value of the numerator at $x = c$ is $e^{-c} \neq 0$, while the value of the denominator is 0 at $x = c$. Thus, $\lim_{x \rightarrow c^+} \frac{e^{-x}}{x-c} = \infty$ and $\lim_{x \rightarrow c^-} \frac{e^{-x}}{x-c} = -\infty$. This makes the line $x = c$ a vertical asymptote for q .
- b. We can write $q(x)$ as $q(x) = \frac{1}{(x-c)e^x}$ and thus see the denominator increases without bound as $x \rightarrow \infty$. Therefore, $\lim_{x \rightarrow \infty} \frac{e^{-x}}{x-c} = 0$. This makes the line $y = 0$ a horizontal asymptote for q . As $x \rightarrow -\infty$, e^{-x} increases without bound, while $(x-c)$ decreases without bound. Thus, q has an indeterminate form as $x \rightarrow -\infty$. Applying L'Hopital's Rule gives us

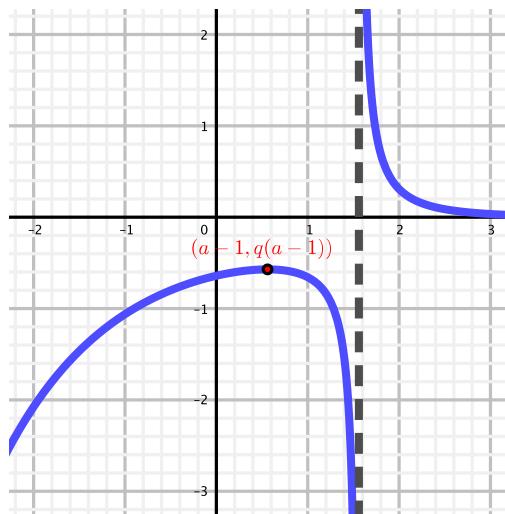
$$\lim_{x \rightarrow -\infty} \frac{e^{-x}}{x-c} = \lim_{x \rightarrow -\infty} \frac{-e^{-x}}{1} = -\infty.$$

- c. Using $q(x) = \frac{1}{(x-c)e^x}$, the quotient and product rules along with algebraic simplification show that

$$q'(x) = -\frac{(x-c)e^x + e^x}{(x-c)^2(e^x)^2} = -\frac{x-c+1}{(x-c)^2e^x}.$$

The function q has critical points in its domain where $q'(x) = 0$ or $q'(x)$ doesn't exist. Note that $q'(x)$ is undefined at $x = c$, but c is not in the domain of q . Thus, the only critical points of q occur when $q'(x) = 0$. The only way the fraction $q'(x)$ can be 0 is if the numerator is 0. Since $x - c + 1 = 0$ when $x = c - 1$, the only critical number for q is $x = c - 1$.

- d. The values of the exponential function e^x and the quadratic $(x - c)^2$ are positive for $x \neq c$, so the sign of $q'(x) = -\frac{x-c+1}{(x-c)^2 e^x}$ depends entirely on the sign of $x - (c - 1)$. When $x < c - 1$ we see that $q'(x) > 0$ and when $x > c - 1$ we have $q'(x) < 0$. Thus, q is increasing on $(-\infty, c - 1)$ and is decreasing on $(c - 1, c)$, so it follows that $q(c - 1)$ is a relative maximum value. This value is a local maximum, since the value of q can be made as large as we like as shown in part (b).
- e. We've shown that a typical member of the family of functions given by $q(x) = \frac{1}{(x-c)e^x}$ has a vertical asymptote at $x = c$ and a critical number at which a local maximum occurs one unit to the left of c at $x = c - 1$. A typical such graph is shown below. Note how q is increasing and decreasing according to our earlier work in (d).



5. Let $E(x) = e^{-\frac{(x-m)^2}{2s^2}}$, where m is any real number and s is a positive real number.
- Compute $E'(x)$ and hence find all critical numbers of E .
 - Construct a first derivative sign chart for E and classify each critical number of the function as a local minimum, local maximum, or neither.
 - It can be shown that $E''(x)$ is given by the formula

$$E''(x) = e^{-\frac{(x-m)^2}{2s^2}} \left(\frac{(x-m)^2 - s^2}{s^4} \right).$$

Find all values of x for which $E''(x) = 0$.

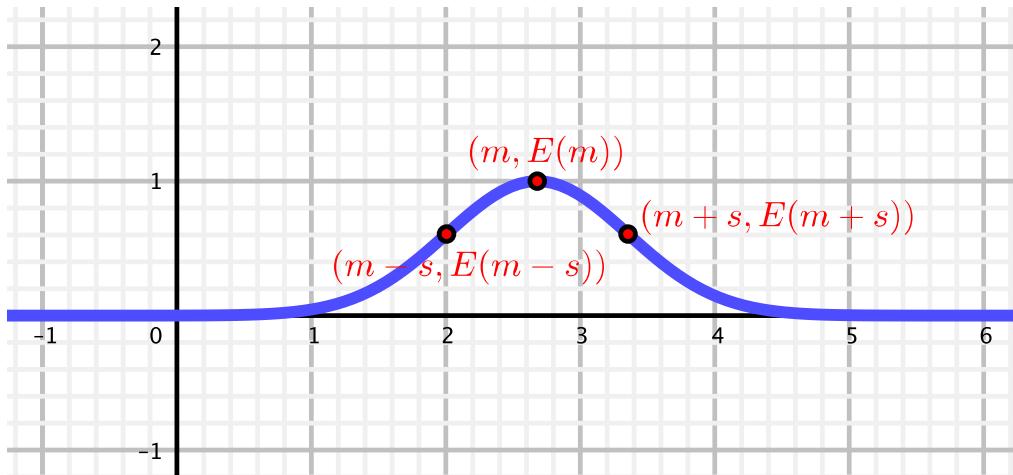
- Determine $\lim_{x \rightarrow \infty} E(x)$ and $\lim_{x \rightarrow -\infty} E(x)$.
- Construct a labeled graph of a typical function E that clearly shows how important points on the graph of $y = E(x)$ depend on m and s .

Answer.

- $x = m$.
- E is increasing for $x < m$ and decreasing for $x > m$, with a local maximum at $x = m$.
- $x = m \pm s$.

d. $\lim_{x \rightarrow \infty} E(x) = \lim_{x \rightarrow -\infty} E(x) = 0$.

e.



Solution.

a. Treating m and s as constants, by the chain rule,

$$E'(x) = e^{-\frac{(x-m)^2}{2s^2}} \cdot \frac{d}{dx} \left[-\frac{(x-m)^2}{2s^2} \right] = e^{-\frac{(x-m)^2}{2s^2}} \cdot \left[-\frac{2(x-m)}{2s^2} \right]$$

Note that $E'(x)$ is defined for every value of x , thus the only critical numbers occur where $E'(x) = 0$. Since e raised to any power is never zero, the only way that $E'(x) = 0$ is for $\left[-\frac{2(x-m)}{2s^2} \right] = 0$, which implies that $x = m$. Thus, $x = m$ is the only critical number of E .

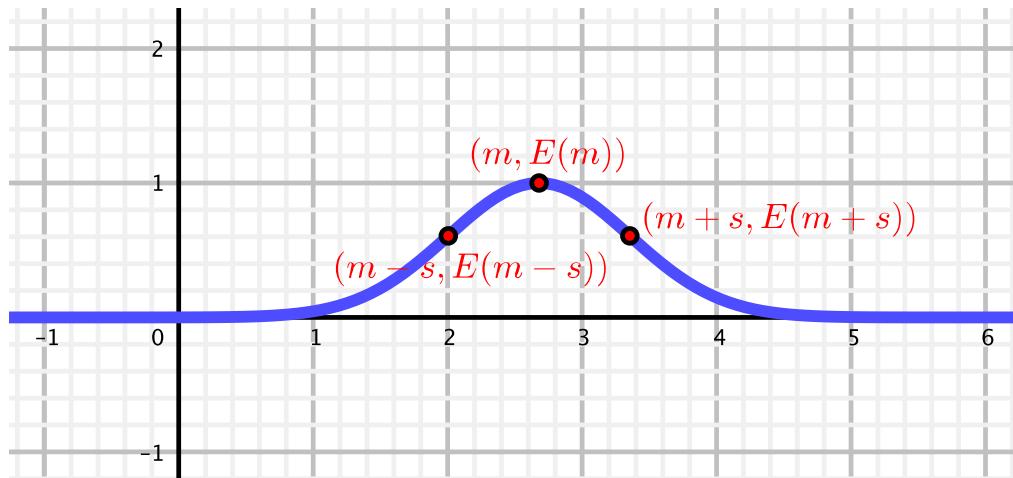
- b. As we noted in (a), $x = m$ is the only critical number of E . Moreover, since e to any power is always positive, the sign of $E'(x)$ depends entirely on the expression $-\frac{2(x-m)}{2s^2} = -\frac{1}{2}(x-m)$. This linear function of x has negative slope, and thus is positive for $x < m$ and negative for $x > m$, and thus $E'(x)$ is positive for $x < m$ and negative for $x > m$. Therefore, E is increasing for $x < m$ and decreasing for $x > m$, with a local maximum at $x = m$.
- c. From the given formula for $E''(x)$ and our knowledge that e to any power is positive, the only way $E''(x) = 0$ is for

$$\left(\frac{(x-m)^2 - s^2}{s^4} \right) = 0,$$

which implies that $(x-m)^2 = s^2$, and therefore $x - m = \pm s$, or $x = m \pm s$. These are the two values for which $E''(x) = 0$.

- d. As $x \rightarrow \pm\infty$, we know that the quadratic function $-\frac{(x-m)^2}{2s^2}$ must decrease without bound because the quadratic function opens down. Thus, the power to which we are raising e in the formula for $E(x)$ is decreasing without bound, which means that $E(x)$ must approach 0 since raising e to more and more negative numbers drives the result to 0. Thus, $\lim_{x \rightarrow \infty} E(x) = \lim_{x \rightarrow -\infty} E(x) = 0$.

- e. We've shown that E has a single critical number at $x = m$ and two potential points of inflection at $x = m \pm s$, as well as that $E(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. All of these properties are reflected in the following figure.



3.5 Related Rates

Motivating 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?
-

In most of our applications of the derivative so far, we have been interested in the instantaneous rate at which one variable, say y , changes with respect to another, say x , leading us to compute and interpret $\frac{dy}{dx}$. We next consider situations where several variable quantities are related, but where each quantity is implicitly a function of time, which will be represented by the variable t . 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 so that its volume increases at a constant rate of 20 cubic inches per second. Since the balloon's volume and radius are related, by knowing how fast the volume is changing, we ought to be able to discover how fast the radius is changing. We are interested in questions such as: can we determine how fast the radius of the balloon is increasing at the moment the balloon's diameter is 12 inches?

Example 3.5.1. 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?

- Draw several spheres with different radii, and observe that as volume changes, the radius, diameter, and surface area of the balloon also change.
- 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 $\frac{dr}{dt}$.
- 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?
- 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 $d = 12$?
- How is the situation different when $d = 16$? When is the radius changing more rapidly, when $d = 12$ or when $d = 16$?

3.5.1 Related Rates Problems

In problems where two or more quantities can be related to one another, and all of the variables involved are implicitly functions of time, t , we are often interested in how their rates are related; we call these *related rates* problems. Once we have an equation establishing the relationship among the variables, we differentiate implicitly with respect to time to find connections among the rates of change.

Example 3.5.2. Sand is being dumped by a conveyor belt onto a pile so that the sand forms a right circular cone, as pictured in Figure 3.5.3.

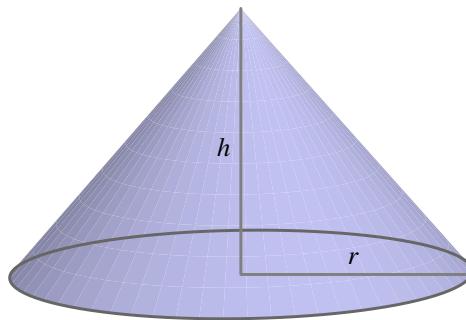


Figure 3.5.3: A conical pile of sand.

Solution. As sand falls from the conveyor belt, 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.

We begin by identifying which variables are changing and how they are related. In this problem, we observe that the radius and height of the pile are related to its volume 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 differentiate implicitly to arrive at an equation that relates their respective rates of change. Taking the derivative of each side of the equation with respect to t , we find

$$\frac{d}{dt}[V] = \frac{d}{dt} \left[\frac{1}{3}\pi r^2 h \right].$$

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 need the product and chain rules. We find that

$$\frac{dV}{dt} = \frac{d}{dt} \left[\frac{1}{3}\pi r^2 h \right]$$

$$\begin{aligned}
 &= \frac{1}{3}\pi r^2 \frac{d}{dt}[h] + \frac{1}{3}\pi h \frac{d}{dt}[r^2] \\
 &= \frac{1}{3}\pi r^2 \frac{dh}{dt} + \frac{1}{3}\pi h 2r \frac{dr}{dt}
 \end{aligned}$$

(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 same principles are applied here when we compute $\frac{d}{dt}[r^2] = 2r \frac{dr}{dt}$.)

The equation

$$\frac{dV}{dt} = \frac{1}{3}\pi r^2 \frac{dh}{dt} + \frac{2}{3}\pi r h \frac{dr}{dt},$$

relates the rates of change of V , h , and r .

If we are given sufficient additional information, we may then find the value of one or more of these rates of change at a specific point in time.

Example 3.5.4. In the setting of Example 3.5.2, suppose we also 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. How fast is the height of the sandpile changing at the moment the radius is 4 feet?

Solution. 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 $\frac{dV}{dt}$ 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

$$\frac{dV}{dt} = \frac{1}{3}\pi r^2 \cdot \frac{1}{2}\frac{dr}{dt} + \frac{2}{3}\pi r \cdot \frac{1}{2}r \cdot \frac{dr}{dt}. \quad (3.5.1)$$

Since sand falls from the conveyor at the constant rate of 10 cubic feet per minute, the value of $\frac{dV}{dt}$, the rate at which the volume of the sand pile changes, is $\frac{dV}{dt} = 10 \text{ ft}^3/\text{min}$. We are interested in how fast the height of the pile is changing at the instant when $r = 4$, so we substitute $r = 4$ and $\frac{dV}{dt} = 10$ into Equation (3.5.1), to find

$$10 = \frac{1}{3}\pi 4^2 \cdot \frac{1}{2}\frac{dr}{dt} \bigg|_{r=4} + \frac{2}{3}\pi 4 \cdot \frac{1}{2}4 \cdot \frac{dr}{dt} \bigg|_{r=4} = \frac{8}{3}\pi \frac{dr}{dt} \bigg|_{r=4} + \frac{16}{3}\pi \frac{dr}{dt} \bigg|_{r=4}.$$

Only the value of $\frac{dr}{dt} \bigg|_{r=4}$ remains unknown. We combine like terms on the right side of the equation above to get $10 = 8\pi \frac{dr}{dt} \bigg|_{r=4}$, and solve for $\frac{dr}{dt} \bigg|_{r=4}$ to find

$$\frac{dr}{dt} \bigg|_{r=4} = \frac{10}{8\pi} \approx 0.39789$$

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

$$\frac{dh}{dt} \bigg|_{r=4} = \frac{5}{8\pi} \approx 0.19894 \text{ ft/min.}$$

Note the difference between the notations $\frac{dr}{dt}$ and $\frac{dr}{dt}|_{r=4}$. 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, the moment when $r = 4$.

Had we known that $h = \frac{1}{2}r$ at the beginning of Example 3.5.2, we could have immediately simplified our work by writing 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

$$\frac{dV}{dt} = \frac{1}{2}\pi r^2 \frac{dr}{dt},$$

from which the same conclusions can be made.

Our work with the sandpile problem above is similar in many ways to our approach in Example 3.5.1, and these steps are typical of most related rates problems. In certain ways, they also resemble work we do in applied optimization problems, and here we summarize the main approach for consideration in subsequent problems.

Note 3.5.5.

- 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 $\frac{dr}{dt}|_{r=4}$.

When identifying variables and drawing a picture, it is important to think about the dynamic ways in which the quantities change. Sometimes a sequence of pictures can be helpful; for some 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>;
- 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>;

¹We again refer to the work of Prof. Marc Renault of Shippensburg University, found at <http://gvsu.edu/s/5p>.

- 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 Example 3.5.6);
- how a skateboarder's shadow changes as he moves past a lamppost <http://gvsu.edu/s/9q>.

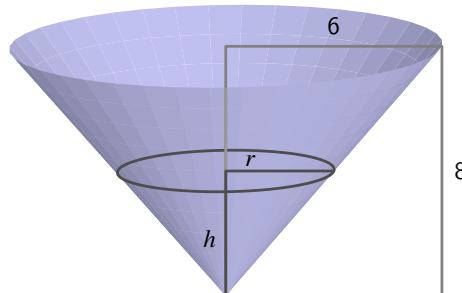
Drawing well-labeled diagrams and envisioning how different parts of the figure change is a key part of understanding related rates problems and being successful at solving them.

Example 3.5.6. 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.

- 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.
- 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?
- 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.
- 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 .
- Find the instantaneous rate at which the water level is rising when the water in the tank is 3 feet deep.
- When is the water rising most rapidly: at $h = 3$, $h = 4$, or $h = 5$?

Hint.

a.

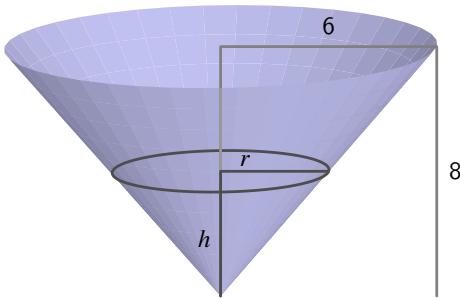


- Think about similar triangles.
- Recall that the volume of a cone is $V = \frac{1}{3}\pi r^2 h$.
- Remember to differentiate implicitly with respect to t .

- e. Use $h = 3$ and the fact that the value of $\frac{dV}{dt}$ is given.
- f. Consider <http://gvsu.edu/s/9p>. Why does this phenomenon occur?

Answer.

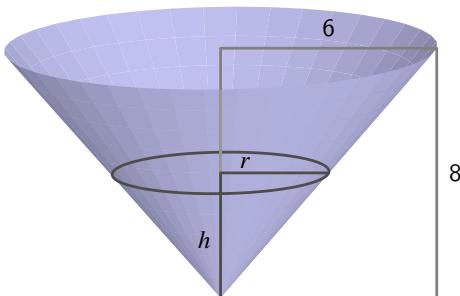
a.



- b. $r = \frac{3}{4}h$.
- c. $V = \frac{3}{16}\pi h^3$.
- d. $\frac{dV}{dt} = \frac{9}{16}\pi h^2 \frac{dh}{dt}$.
- e. $\frac{dh}{dt} \Big|_{h=3} = \frac{64}{81\pi} \approx 0.2515$ feet per minute.

Solution.

- a. Letting r represent the water's radius at time t and h the water's depth, we see the following situation:



- b. Observe that the right triangle with legs of length h and r is similar to the right triangle with legs of length 8 and 6, respectively, based on how the water assumes the shape of the tank, and thus $\frac{r}{h} = \frac{6}{8}$, so that $r = \frac{3}{4}h$.

- c. Since the water in the tank always takes the shape of a circular cone, the volume of water in the tank at time t is given by $V = \frac{1}{3}\pi r^2 h$. Because we have established that $r = \frac{3}{4}h$, it follows that

$$V = \frac{1}{3}\pi \left(\frac{3}{4}h\right)^2 h = \frac{3}{16}\pi h^3.$$

- d. Differentiating with respect to t , we now find

$$\frac{dV}{dt} = \frac{9}{16}\pi h^2 \frac{dh}{dt},$$

which relates the rates of change of V and h .

- e. It is given in the problem setting that water is entering the tank at a rate of 4 cubic feet per minute, hence $\frac{dV}{dt} = 4$, and we are interested in the rate of change of the water's depth when $h = 3$. Substituting these values into the equation that relates $\frac{dV}{dt}$ and $\frac{dh}{dt}$, we find that

$$4 = \frac{9}{16}\pi 3^2 \frac{dh}{dt} \Big|_{h=3},$$

so that $\frac{dh}{dt} \Big|_{h=3} = \frac{64}{81\pi} \approx 0.2515$ feet per minute.

Recognizing which geometric relationships are relevant in a given problem is often the key to finding the function to optimize. For instance, although the problem in Example 3.5.6 is about a conical tank, the most important fact is that there are two similar right triangles involved. In another setting, we might use the Pythagorean Theorem to relate the legs of the triangle. But in the conical tank, the fact that the water fills the tank so that that the ratio of radius to depth is constant turns out to be the important relationship. In other situations where a changing angle is involved, trigonometric functions may provide the means to find relationships among various parts of the triangle.

Example 3.5.7. 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.)

- 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.
- 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).
- 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).

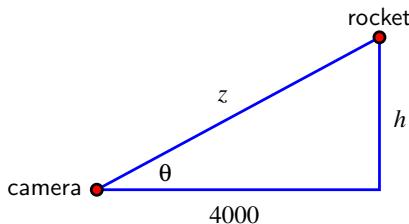
- d. Suppose that the rocket's speed is 600 ft/sec at the instant it has risen 3000 feet. How fast is the distance from the television camera to the rocket changing at that moment? If the camera is following the rocket, how fast is the camera's angle of elevation changing at that same moment?
- e. If 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?

Hint.

- Let θ represent the camera angle and note that one leg of the right triangle is constant. Which two are changing?
- Think trigonometrically.
- Think like Pythagoras.
- Use the facts that $h = 3000$ and $\frac{dh}{dt} = 600$ in your preceding work.
- You can answer this question intuitively or by changing the value of h in your work in (d).

Answer.

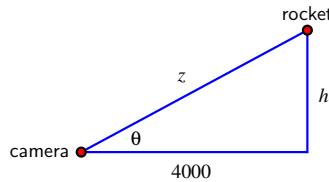
a.



- $\frac{dh}{dt} = 4000 \sec^2(\theta) \frac{d\theta}{dt}$.
- $h \frac{dh}{dt} = z \frac{dz}{dt}$.
- $\frac{dz}{dt} \big|_{h=3000} = 360$ feet/sec; $\frac{d\theta}{dt} \big|_{h=3000} = \frac{12}{125}$ radians per second.
- greater.

Solution.

- a. Letting θ be the camera's elevation angle, h the rocket's height, and z the distance from the camera to the rocket, we have the following situation at a given point in time:



- b. To relate θ and h , observe that the tangent function is a good choice, since $\tan(\theta) = \frac{h}{4000}$, so that

$$h = 4000 \tan(\theta).$$

Differentiating implicitly with respect to t , we find that

$$\frac{dh}{dt} = 4000 \sec^2(\theta) \frac{d\theta}{dt}.$$

- c. To relate z and h , the Pythagorean Theorem is natural to consider. By this famous result,

$$h^2 + 4000^2 = z^2.$$

Differentiating both sides implicitly with respect to t , it follows

$$2h \frac{dh}{dt} = 2z \frac{dz}{dt},$$

and thus $h \frac{dh}{dt} = z \frac{dz}{dt}$.

- d. Using the given fact that the rocket's speed is 600 ft/sec at the instant it has risen 3000 feet, we know that $\frac{dh}{dt} \Big|_{h=3000} = 600$. Note further in the triangle that when $h = 3000$, it follows $z = 5000$, since the base leg of the triangle is constant at 4000, by using the Pythagorean Theorem. Substituting this information from the instant $h = 3000$ into the equation that relates the rates of change of z and h found in (c), we find that

$$2 \cdot 3000 \cdot 600 = 2 \cdot 5000 \cdot \frac{dz}{dt} \Big|_{h=3000}.$$

Solving for $\frac{dz}{dt} \Big|_{h=3000}$ we have

$$\frac{dz}{dt} \Big|_{h=3000} = \frac{1800}{5} = 360 \text{ feet/sec.}$$

To answer the question about how fast the camera angle is changing, we use the same information but now in the equation we found in (b) that relates the rates of change of θ and h :

$$\frac{dh}{dt} = 4000 \sec^2(\theta) \frac{d\theta}{dt}.$$

Observe that when $h = 3000$, in the 3000-4000-5000 right triangle, it follows that $\cos(\theta) = \frac{4}{5}$, so $\sec(\theta) = \frac{5}{4}$. Thus, using the instantaneous information,

$$600 = 4000 \cdot \frac{25}{16} \frac{d\theta}{dt} \bigg|_{h=3000},$$

and thus

$$\frac{d\theta}{dt} \bigg|_{h=3000} = \frac{6 \cdot 16}{40 \cdot 25} = \frac{12}{125},$$

which is measured in radians per second.

- e. Recalling that $2h \frac{dh}{dt} = 2z \frac{dz}{dt}$, it follows that

$$\frac{dz}{dt} = \frac{h}{z} \frac{dh}{dt}.$$

Using this equation when $h = 3000$ and $\frac{dh}{dt} = 600$ led us to conclude that $\frac{dz}{dt} \bigg|_{h=3000} = \frac{3}{5} \cdot 600 = 360$ feet/sec. If we instead use $h = 4000$, it follows that $z = 4000\sqrt{2}$, so that

$$\frac{dz}{dt} \bigg|_{h=4000} = \frac{4}{4\sqrt{2}} \cdot 600 \approx 424.26 \text{ feet/sec.}$$

Indeed, $\frac{dz}{dt}$ is an increasing function of h , provided that $\frac{dh}{dt}$ is constant, because we can write $h^2 + 4000^2 = z^2$, so $z = \sqrt{h^2 + 4000^2}$, making

$$\frac{dz}{dt} = \frac{h}{\sqrt{h^2 + 4000^2}} \frac{dh}{dt} = \frac{h}{\sqrt{h^2 + 4000^2}} 600.$$

It is straightforward to verify that $\frac{h}{\sqrt{h^2 + 4000^2}}$ is an increasing function of h .

In addition to finding instantaneous rates of change at particular points in time, we can often make more general observations about how particular rates themselves will change over time. For instance, when a conical tank is filling with water at a constant rate, it seems obvious that the depth of the water should increase more slowly over time. Note how carefully we must phrase the relationship: 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). We make this observation by solving the equation that relates the various rates for one particular rate, without substituting any particular values for known variables or rates. For instance, in the conical tank problem in Example 3.5.6, 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 but remain positive. Hence, the depth of the water is increasing at a decreasing rate.

Example 3.5.8. 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.

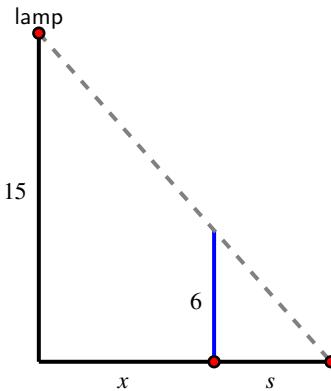
- Draw an appropriate right triangle that represents a snapshot in time of the skateboarder, lamppost, and his shadow. Let x denote the horizontal distance from the base of the lamppost to the skateboarder and s represent the length of his shadow. Label these quantities, as well as the skateboarder's height and the lamppost's height on the diagram.
- 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 .
- Use your work in (b) to find an equation that relates $\frac{dx}{dt}$ and $\frac{ds}{dt}$.
- At what rate is the length of the skateboarder's shadow increasing at the instant the skateboarder is 8 feet from the lamppost?
- 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?
- Which is moving more rapidly: the skateboarder or the tip of his shadow? Explain, and justify your answer.

Hint.

- Note that the lengths of the legs of the right triangle will be 15 for the vertical one and $x + s$ for the horizontal one.
- The small triangle formed by the skateboarder and his shadow, with legs 6 and s is similar to the large triangle that has the lamppost as one of its legs.
- Simplify the equation in (b) as much as possible before differentiating implicitly with respect to t .
- Find $\frac{ds}{dt}|_{x=8}$.
- Does the equation that relates $\frac{dx}{dt}$ and $\frac{ds}{dt}$ involve x ? Is $\frac{dx}{dt}$ changing or constant?
- Let y represent the location of the tip of the shadow, so that $y = x + s$.

Answer.

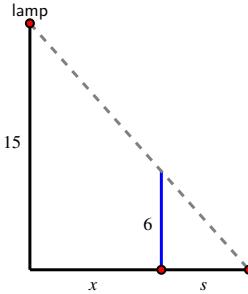
-



- b. $3s = 2x$.
- c. $3\frac{ds}{dt} = 2\frac{dx}{dt}$.
- d. $\frac{ds}{dt}|_{x=8} = 2$ feet per second.
- e. at a constant rate.
- f. Let y represent the location of the tip of the shadow; $\frac{dy}{dt} = 5$ feet/sec.

Solution.

- a. The given information leads us to construct the following diagram:



- b. The small triangle formed by the skateboarder and his shadow, with legs of length 6 and s is similar to the large triangle that has the lamppost as one of its legs (length 15) and horizontal leg of length $x + s$. Because the ratios of the lengths of the legs of these two triangles is equal, we have

$$\frac{s}{6} = \frac{x+s}{15}.$$

Simplifying, we have $15s = 6s + 6x$, so that $9s = 2x$, or most simply, $3s = 2x$.

- c. Differentiating with respect to t , it is immediate that $3\frac{ds}{dt} = 2\frac{dx}{dt}$.
- d. Since $\frac{ds}{dt} = \frac{2}{3}\frac{dx}{dt}$, and $\frac{dx}{dt} = 3$, it follows $\frac{ds}{dt} = 2$ for every value of t (and x). Thus, $\frac{ds}{dt}|_{x=8} = 2$ feet per second.

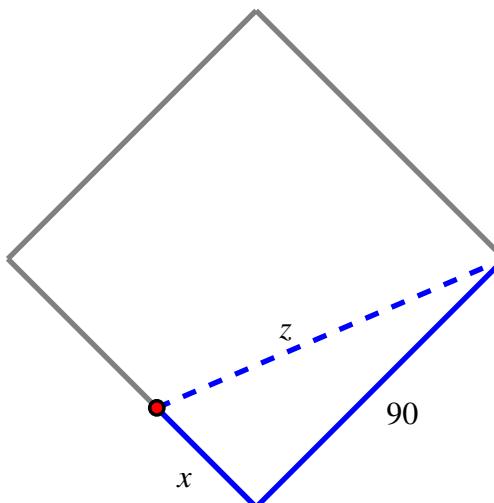
- e. Because $\frac{ds}{dt}$ is constant, the shadow's length is increasing at a constant rate (irrespective of the distance from the skateboarder to the lamppost).
- f. Let y represent the location of the tip of the shadow, so that $y = x + s$. Observe that we can now compute $\frac{dy}{dt}$ in terms of $\frac{dx}{dt}$ and $\frac{ds}{dt}$, with $\frac{dy}{dt} = \frac{dx}{dt} + \frac{ds}{dt} = 3 + 2 = 5$ feet/sec, and hence the tip of the shadow is moving more rapidly than the skateboarder himself.

In the first three activities of this section, we provided guided instruction to build a solution in a step by step way. For the closing example and the following exercises, most of the detailed work is left to the reader.

Example 3.5.9. A baseball diamond is 90' square. A batter hits a ball along the third base line and 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?

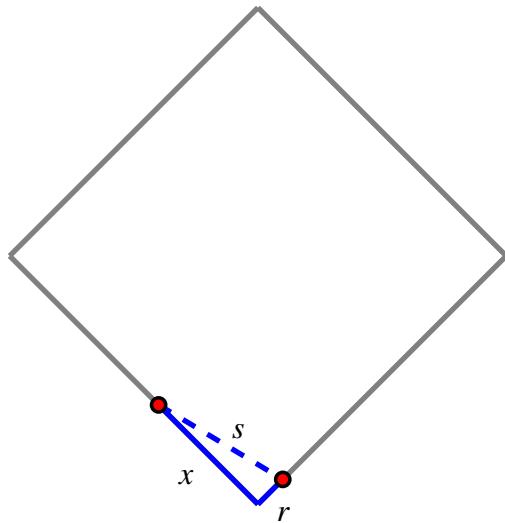
Hint. Let x denote the position of the ball along the third base line at time t , and z the distance from the ball to first base. Note that the basepaths meet at 90 degree angles.

Answer. Let x denote the position of the ball at time t and z the distance from the ball to first base, as pictured below.



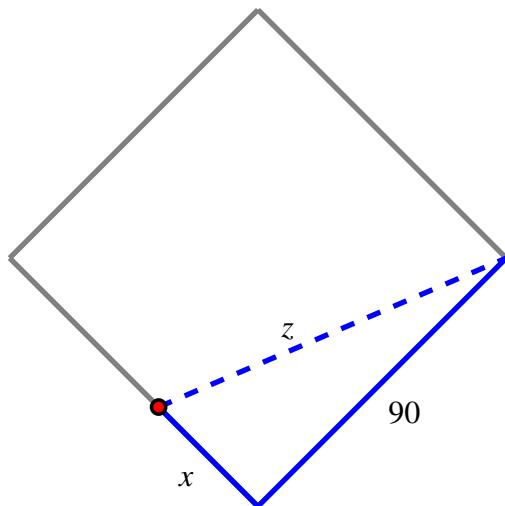
$$\frac{dz}{dt} \Big|_{x=45} = \frac{100}{\sqrt{5}} \approx 44.7214 \text{ feet/sec.}$$

Let r be the runner's position at time t and let s be the distance between the runner and the ball, as pictured.



$$\frac{ds}{dt} \Big|_{x=45} = \frac{430}{\sqrt{17}} \approx 104.2903 \text{ feet/sec.}$$

Solution. We let x denote the position of the ball at time t and z the distance from the ball to first base, as pictured below.



By the Pythagorean Theorem, we know that $x^2 + 90^2 = z^2$; differentiating with respect to t , we have

$$2x \frac{dx}{dt} = 2z \frac{dz}{dt}.$$

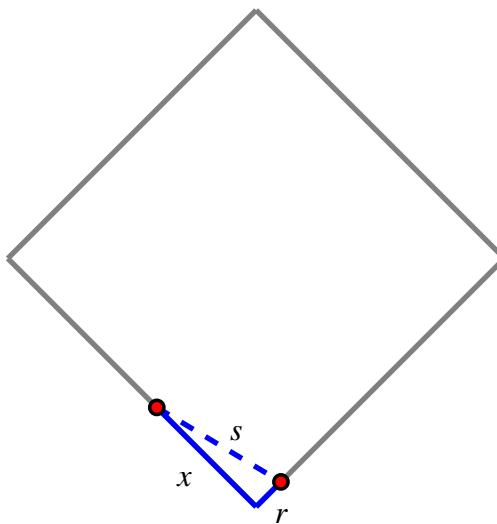
At the instant the ball is halfway to third base, we know $x = 45$ and $\frac{dx}{dt} \Big|_{x=45} = 100$. Moreover, by Pythagoras, $z^2 = 90^2 + 45^2$, so $z = 45\sqrt{5}$. Thus,

$$2 \cdot 45 \cdot 100 = 2 \cdot 45\sqrt{5} \cdot \frac{dz}{dt} \Big|_{x=45},$$

so

$$\left. \frac{dz}{dt} \right|_{x=45} = \frac{100}{\sqrt{5}} \approx 44.7214 \text{ feet/sec.}$$

For the second question, we still let x represent the ball's position at time t , but now we introduce r as the runner's position at time t and let s be the distance between the runner and the ball. In this setting, as seen in the diagram below,



x , r , and s form the sides of a right triangle, so that

$$x^2 + r^2 = s^2,$$

by the Pythagorean Theorem. Differentiating each side with respect to t , it follows that the three rates of change are related by the equation

$$2x \frac{dx}{dt} + 2r \frac{dr}{dt} = 2s \frac{ds}{dt}.$$

We are given that at the instant $x = 45$, $r = \frac{90}{8}$, so by Pythagoras, $s = \frac{45}{4}\sqrt{17}$. In addition, at this same instant we know that $\left. \frac{dx}{dt} \right|_{x=45} = 100$ and $\left. \frac{dr}{dt} \right|_{x=45} = 30$. Applying this information,

$$2 \cdot 45 \cdot 100 + 2 \cdot \frac{45}{4} \cdot 30 = 2 \cdot \frac{45}{4}\sqrt{17} \cdot \left. \frac{ds}{dt} \right|_{x=45}$$

and therefore

$$\left. \frac{ds}{dt} \right|_{x=45} = \frac{430}{\sqrt{17}} \approx 104.2903 \text{ feet/sec.}$$

3.5.2 Summary

- 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

$$2x \frac{dx}{dt} + 2y \frac{dy}{dt} = 2z \frac{dz}{dt},$$

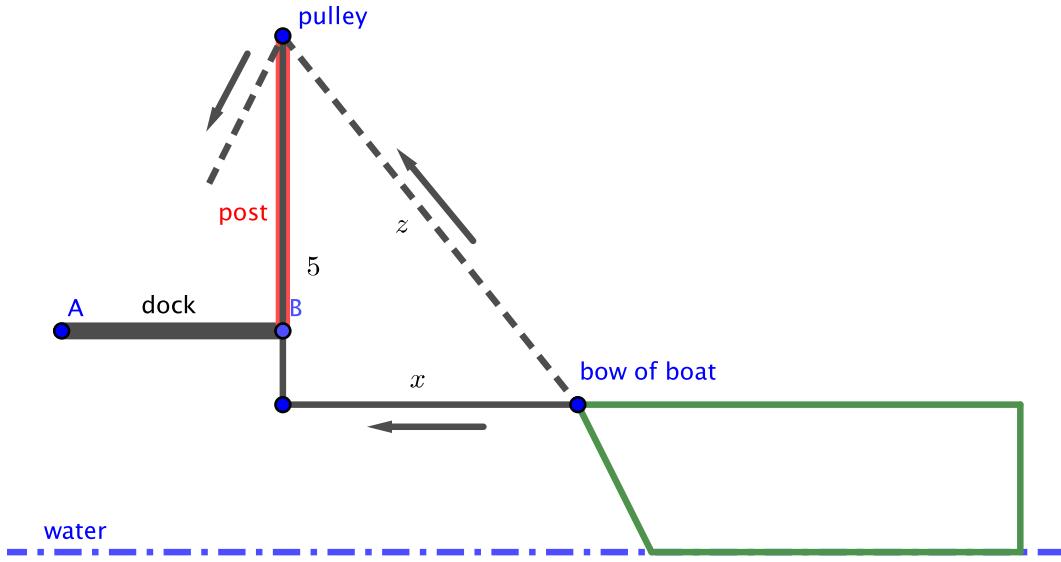
so that if we know the values of x , y , and z at a particular time, as well as two of the three rates, we can deduce the value of the third.

3.5.3 Exercises

- (Height of a conical pile of gravel) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Movement of a shadow) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (A leaking conical tank) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- A sailboat is sitting at rest near its dock. A rope attached to the bow of the boat is drawn in over a pulley that stands on a post on the end of the dock that is 5 feet higher than the bow. If the rope is being pulled in at a rate of 2 feet per second, how fast is the boat approaching the dock when the length of rope from bow to pulley is 13 feet?

Answer. The boat is approaching the dock at a rate of $\frac{13}{6} \approx 2.167$ feet per second.

Solution. Using the given information, we construct the figure shown below.



As pictured, we know that 5 is the vertical height from the pulley to the level of the bow of the boat, and we let z be the length of the rope from the pulley to the bow of the boat, and x the horizontal distance from the dock to the bow of the boat.

We are given that the rope is being pulled in at 2 feet per second, and thus $\frac{dz}{dt} = -2$ feet per second. Because we want to know how fast the boat is approaching the dock when the length of rope from bow to pulley is 13 feet, we want to know $\frac{dx}{dt}|_{z=13}$. Thus, we need to relate the changing quantities z and x .

Because the rope, the post (vertically extended), and the horizontal distance from the bow of the boat to the post on the dock form a right triangle at all times, it follows that

$$x^2 + 5^2 = z^2.$$

Having now related z and x , we differentiate this equation with respect to t . By the chain rule, we now see that

$$2x \frac{dx}{dt} = 2z \frac{dz}{dt}.$$

At the instant $z = 13$, $x^2 + 5^2 = 13^2$, and thus $x = 12$. Using all of the given information at the instant $z = 13$ (including that $\frac{dz}{dt} = -2$),

$$2(12) \frac{dx}{dt} \Big|_{z=13} = 2(13)(-2).$$

Solving for $\frac{dx}{dt}|_{z=13}$,

$$\frac{dx}{dt} \Big|_{z=13} = -\frac{13}{6}$$

feet per second. Thus the boat is approaching the dock at a rate of $\frac{13}{6} \approx 2.167$ feet per second.

5. 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 in the Figure 3.5.10.

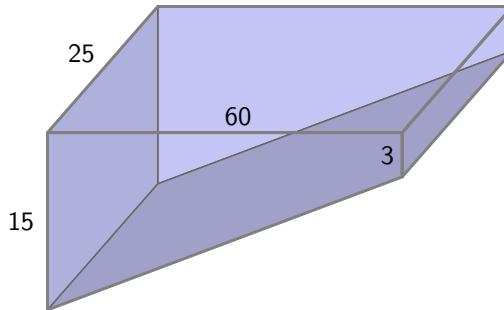


Figure 3.5.10: The swimming pool.

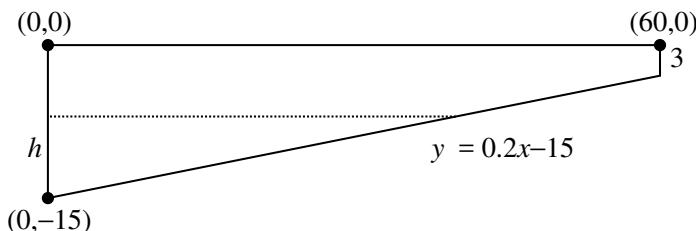
Suppose the pool has been emptied and is now being filled with water at a rate of 800 cubic feet per minute. At what rate is the depth of water (measured at the deepest point of the pool) increasing when it is 5 feet deep at that end? Over time, describe how the depth of the water will increase: at an increasing rate, at a decreasing rate, or at a constant rate. Explain.

Answer. The depth of the water is increasing at

$$\left. \frac{dh}{dt} \right|_{h=5} = 1.28$$

feet per minute. The depth of the water is increasing at a decreasing rate.

Solution. The variables in this problem are the volume V of water in the pool and the depth y of water in the pool (measured at the deepest point of the pool); each is implicitly a function of time, t . We are given that $\frac{dV}{dt}$ is a constant 800 cubic feet per minute and want to find $\left. \frac{dh}{dt} \right|_{h=5}$. To do this, we first need to relate V and h . The volume of the water in the pool at height $h < 12$ is the volume of the triangular cross sectional area of height h (as shown in the figure below) times the width (25 feet) of the pool. The height of the triangular cross section is h and the length is distance of the dotted line indicated in the figure.



Placing the cross section on a coordinate system as in the figure, the hypotenuse of the triangle is the line connecting the points $(0, -15)$ and $(60, -3)$. This line has slope $\frac{12}{60} = 0.2$ and y -intercept $(0, -15)$. So the equation of this line is $y = 0.2x - 15$. It is the x coordinate of the point on this line corresponding to the y -coordinate $h - 15$ that is the length of the triangle whose height is h . So the length l is $l = \frac{(h-15)+15}{0.2} = 5h$. Thus, the volume of water in the pool

at height h is

$$V = \left(\frac{1}{2}(h)(5h)\right)(25) = 62.5h^2.$$

Both V and h are functions of time. Differentiating both sides of the equation with respect to t gives

$$\frac{dV}{dt} = 62.5(2h)\frac{dh}{dt},$$

so

$$\frac{dh}{dt} = \frac{1}{125h} \frac{dV}{dt}.$$

When the depth of water in the pool is 5 feet, then the depth is increasing at the rate

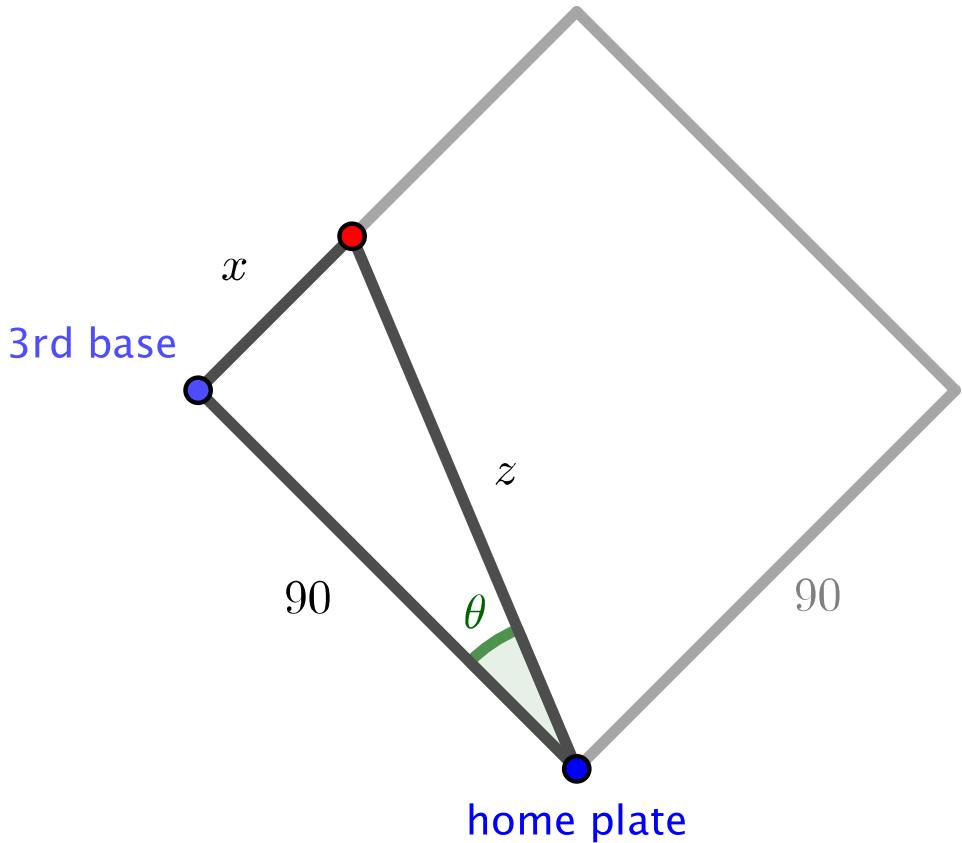
$$\frac{dh}{dt} \Big|_{h=5} = \frac{1}{125(5)}(800) = 1.28$$

feet per minute. Since $\frac{dh}{dt}$ is inversely proportional to h , the rate at which the depth of the water increases slows as h increases, so the depth of the water is increasing at a decreasing rate.

6. 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?

Answer. $\frac{d\theta}{dt}|_{x=30} = -0.24$ radians per second.

Solution. Let x represent the distance from the runner to third base along the path from second to third, let z be the distance from the umpire at home plate to the runner, and call θ the angle between the third baseline and the line of sight from the umpire to the runner. See the figure below.



Each of x , z , and θ is implicitly a function of t . We are given that the runner is advancing at a rate of 24 feet per second, so $\frac{dx}{dt} = -24$. Since we want to know $\frac{d\theta}{dt}|_{x=30}$, we want to relate x and θ .

We first observe that by definition,

$$\tan(\theta) = \frac{x}{90} = \frac{1}{90}x.$$

Having related x and θ , we can differentiate both sides of the preceding equation with respect to t . By the chain rule,

$$\sec^2(\theta) \frac{d\theta}{dt} = \frac{1}{90} \frac{dx}{dt},$$

and thus

$$\frac{d\theta}{dt} = \frac{1}{90 \sec^2(\theta)} \frac{dx}{dt}$$

At the instant $x = 30$, since $x^2 + 90^2 = z^2$, $z = \sqrt{90^2 + 30^2} = \sqrt{9000} = 30\sqrt{10}$. Hence, at this same instance, $\sec(\theta) = \frac{1}{\cos(\theta)} = \frac{30\sqrt{10}}{90} = \frac{\sqrt{10}}{3}$. Recall we also know that $\frac{dx}{dt} = -24$. Applying

this information to the equation that relates the rates of change of x and θ ,

$$\left. \frac{d\theta}{dt} \right|_{x=30} = \frac{1}{90 \left(\frac{\sqrt{10}}{3} \right)^2} (-24) = -\frac{1}{100} (24) = -0.24$$

radians per second.

7. 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?

Answer. $\left. \frac{dh}{dt} \right|_{V=1000} = \frac{10}{\pi \left(\sqrt[3]{\frac{3000}{\pi}} \right)^2} \approx 0.0328$ feet per minute.

Solution. We can draw a figure similar to Figure 3.5.3 and let h be the height of the pile at time t and r the pile's base radius. Given that $\frac{dV}{dt} = 10$ cubic feet per minute, we want to know $\left. \frac{dh}{dt} \right|_{V=1000}$.

We first relate V and h . We know that $V = \frac{1}{3}\pi r^2 h$, as well as that for every time, t , the height and radius are equal so $r = h$. Thus,

$$V = \frac{1}{3}\pi h^2 h = \frac{1}{3}\pi h^3$$

Differentiating both sides of this equation with respect to t

$$\frac{dV}{dt} = \frac{1}{3}\pi \cdot 3h^2 \frac{dh}{dt} = \pi h^2 \frac{dh}{dt}.$$

We note that at the instant $V = 1000$, $1000 = \frac{1}{3}\pi h^3$, so $h^3 = \frac{3000}{\pi}$ and therefore $h = \sqrt[3]{\frac{3000}{\pi}}$.

Recall also that $\frac{dV}{dt} = 10$. Applying all of this information at the instant $V = 1000$ in the equation that relates $\frac{dV}{dt}$ and $\frac{dh}{dt}$, we have

$$10 = \pi \left(\sqrt[3]{\frac{3000}{\pi}} \right)^2 \left. \frac{dh}{dt} \right|_{V=1000}.$$

Thus,

$$\left. \frac{dh}{dt} \right|_{V=1000} = \frac{10}{\pi \left(\sqrt[3]{\frac{3000}{\pi}} \right)^2} \approx 0.0328$$

feet per minute.

3.6 Using Derivatives to Evaluate Limits

Motivating 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 $\lim_{x \rightarrow \infty} f(x) = L$ and $\lim_{x \rightarrow a} f(x) = \infty$?
- How can derivatives assist us in evaluating indeterminate limits of the form $\frac{\infty}{\infty}$?

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 definition of the derivative always generates the indeterminate form $\frac{0}{0}$. If f is a differentiable function, then in the definition

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

not only does $h \rightarrow 0$ in the denominator, but also $(f(x+h) - f(x)) \rightarrow 0$ in the numerator, since f is continuous. Remember, saying that 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 $f'(x) = x^2$.

We have learned many techniques for evaluating the limits that result from the derivative definition, including a large number of shortcut rules. In this section, we turn the situation upside-down: instead of using limits to evaluate derivatives, we explore how to use derivatives to evaluate certain limits.

Example 3.6.1. Let h be the function given by $h(x) = \frac{x^5+x-2}{x^2-1}$.

- What is the domain of h ?
- Explain why $\lim_{x \rightarrow 1} \frac{x^5+x-2}{x^2-1}$ results in an indeterminate form.
- 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.
- Explain why $h(x) \approx \frac{L_f(x)}{L_g(x)}$ for x near $a = 1$.
- Using your work from (c) and (d), evaluate

$$\lim_{x \rightarrow 1} \frac{L_f(x)}{L_g(x)}.$$

What do you think your result tells us about $\lim_{x \rightarrow 1} h(x)$?

- Investigate the function $h(x)$ graphically and numerically near $x = 1$. What do you think is the value of $\lim_{x \rightarrow 1} h(x)$?

3.6.1 Using derivatives to evaluate indeterminate limits of the form $\frac{0}{0}$.

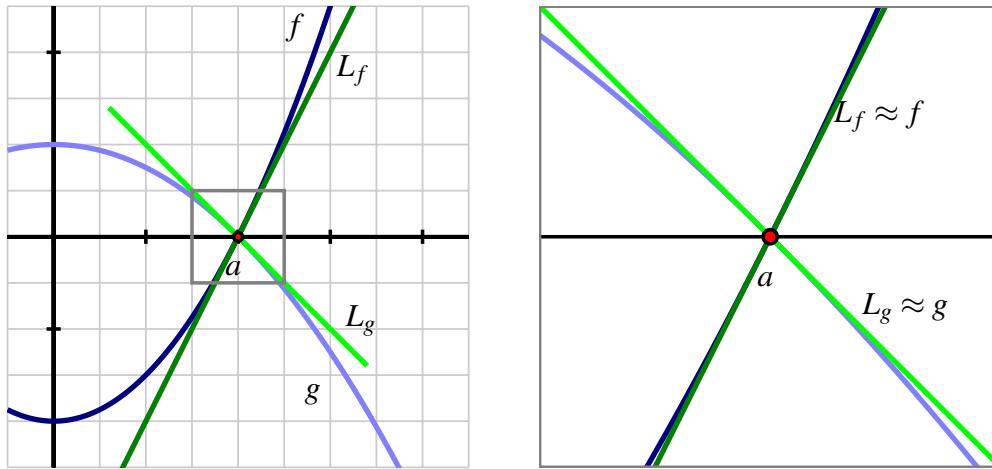


Figure 3.6.2: At left, the graphs of f and g near the value a , along with their tangent line approximations L_f and L_g at $x = a$. At right, zooming in on the point a and the four graphs.

The idea demonstrated in Example 3.6.1 — 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 evaluate a wide range of limits. We will with a function $h(x)$ that can be written as a quotient $h(x) = \frac{f(x)}{g(x)}$, where f and g are both differentiable at $x = a$ and for which $f(a) = g(a) = 0$. We would like to evaluate the indeterminate limit given by $\lim_{x \rightarrow a} h(x)$. Figure 3.6.2 illustrates the situation. We see that both f and g have an x -intercept at $x = a$. Their respective tangent line approximations L_f and L_g at $x = a$ are also shown in the figure. We can take advantage of the fact that a function and its tangent line approximation become indistinguishable as $x \rightarrow a$.

First, let's recall that $L_f(x) = f'(a)(x - a) + f(a)$ and $L_g(x) = g'(a)(x - a) + g(a)$. Because x is getting arbitrarily close to a when we take the limit, we can replace f with L_f and replace g with L_g , and thus we observe that

$$\begin{aligned} \lim_{x \rightarrow a} \frac{f(x)}{g(x)} &= \lim_{x \rightarrow a} \frac{L_f(x)}{L_g(x)} \\ &= \lim_{x \rightarrow a} \frac{f'(a)(x - a) + f(a)}{g'(a)(x - a) + g(a)}. \end{aligned}$$

Next, we remember that both $f(a) = 0$ and $g(a) = 0$, which is precisely what makes the original limit indeterminate. Substituting these values for $f(a)$ and $g(a)$ in the limit above, we now have

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(a)(x - a)}{g'(a)(x - a)}$$

$$= \lim_{x \rightarrow a} \frac{f'(a)}{g'(a)},$$

where the latter equality holds because $\frac{x-a}{x-a} = 1$ when x is approaching (but not equal to) a . Finally, we note that $\frac{f'(a)}{g'(a)}$ is constant with respect to x , and thus

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

This result holds as long as $g'(a)$ is not equal to zero. The formal name of the result is L'Hopital's Rule.

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 $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}$.

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 and the extra assumption that g' is continuous at $x = a$)

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)},$$

provided the righthand limit exists. This form reflects the basic idea of L'Hopital's Rule: if $\frac{f(x)}{g(x)}$ produces an indeterminate limit of form $\frac{0}{0}$ as $x \rightarrow a$, that limit is equivalent to the limit of the quotient of the two functions' derivatives, $\frac{f'(x)}{g'(x)}$.

For example, if we consider the limit from Example 3.6.1,

$$\lim_{x \rightarrow 1} \frac{x^5 + x - 2}{x^2 - 1},$$

by L'Hopital's Rule we have that

$$\lim_{x \rightarrow 1} \frac{x^5 + x - 2}{x^2 - 1} = \lim_{x \rightarrow 1} \frac{5x^4 + 1}{2x} = \frac{6}{2} = 3.$$

By replacing the numerator and denominator with their respective derivatives, we often replace an indeterminate limit with one whose value we can easily determine.

Example 3.6.3. 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. $\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x}$

c. $\lim_{x \rightarrow 1} \frac{2 \ln(x)}{1 - e^{x-1}}$

b. $\lim_{x \rightarrow \pi} \frac{\cos(x)}{x}$

d. $\lim_{x \rightarrow 0} \frac{\sin(x) - x}{\cos(2x) - 1}$

Hint.

- Remember that $\ln(1) = 0$.
- Note that $x \rightarrow \pi$, not $x \rightarrow 0$.
- Observe that $e^{x-1} \rightarrow 1$ as $x \rightarrow 1$.
- If necessary, L'Hopital's Rule can be applied more than once.

Answer.

- $\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = 1$.
- $\lim_{x \rightarrow \pi} \frac{\cos(x)}{x} = -\frac{1}{\pi}$.
- $\lim_{x \rightarrow 1} \frac{2 \ln(x)}{1 - e^{x-1}} = -2$.
- $\lim_{x \rightarrow 0} \frac{\sin(x) - x}{\cos(2x) - 1} = 0$.

Solution.

- As $x \rightarrow 0$, we see that $\ln(1+x) \rightarrow \ln(1) = 0$, thus this limit has an indeterminate form. By L'Hopital's Rule, we have

$$\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = \lim_{x \rightarrow 0} \frac{\frac{1}{1+x}}{1}.$$

As this limit is no longer indeterminate, we may simply allow $x \rightarrow 0$, and thus we find that

$$\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = \frac{\frac{1}{1+0}}{1} = 1.$$

- Observe that

$$\lim_{x \rightarrow \pi} \frac{\cos(x)}{x} = \frac{\cos(\pi)}{\pi} = -\frac{1}{\pi},$$

since this limit is not indeterminate because the function $\frac{\cos(x)}{x}$ is continuous at $x = \pi$.

- Since $\ln(x) \rightarrow 0$ and $e^{x-1} \rightarrow 1$ as $x \rightarrow 0$, this limit is indeterminate with form $\frac{0}{0}$. Hence, by L'Hopital's Rule,

$$\lim_{x \rightarrow 1} \frac{2 \ln(x)}{1 - e^{x-1}} = \lim_{x \rightarrow 1} \frac{\frac{2}{x}}{-e^{x-1}}.$$

The updated limit is not indeterminate, and allowing $x \rightarrow 1$, we find

$$\lim_{x \rightarrow 1} \frac{2 \ln(x)}{1 - e^{x-1}} = \frac{\frac{2}{1}}{-e^0} = -2.$$

d. Since the given limit is indeterminate of form $\frac{0}{0}$, by L'Hopital's Rule we have

$$\lim_{x \rightarrow 0} \frac{\sin(x) - x}{\cos(2x) - 1} = \lim_{x \rightarrow 0} \frac{\cos(x) - 1}{-2 \sin(2x)}.$$

Now, as $x \rightarrow 0$, $\cos(x) \rightarrow 1$ and $\sin(2x) \rightarrow 0$, which makes the latest limit also indeterminate in form $\frac{0}{0}$. Applying L'Hopital's Rule a second time, we now have

$$\lim_{x \rightarrow 0} \frac{\sin(x) - x}{\cos(2x) - 1} = \lim_{x \rightarrow 0} \frac{-\sin(x)}{-4 \cos(2x)}.$$

In the newest limit, we note that $\sin(x) \rightarrow 0$ but $\cos(2x) \rightarrow 1$ as $x \rightarrow 0$, so the numerator is tending to 0 while the denominator is approaching -4. Thus, the value of the limit is determined to be

$$\lim_{x \rightarrow 0} \frac{\sin(x) - x}{\cos(2x) - 1} = 0.$$

While L'Hopital's Rule can be applied in an entirely algebraic way,

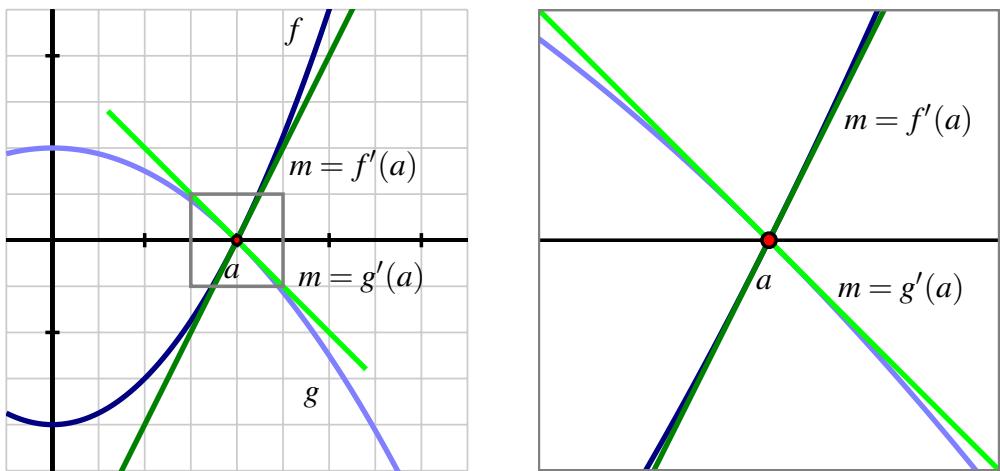


Figure 3.6.4: Two functions f and g that satisfy L'Hopital's Rule.

it is important to remember that the justification 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 $x \rightarrow a$. We see this in Figure 3.6.4, where we can see from the grid that $f'(a) = 2$ and $g'(a) = -1$, hence by L'Hopital's Rule,

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)} = \frac{2}{-1} = -2.$$

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 $f(a) = g(a) = 0$, then the limit of $\frac{f(x)}{g(x)}$ as $x \rightarrow a$ is given by the ratio of the slopes of f and g at $x = a$.

Example 3.6.5. In this example, we reason graphically from the following figure to evaluate limits of ratios of functions about which some information is known.

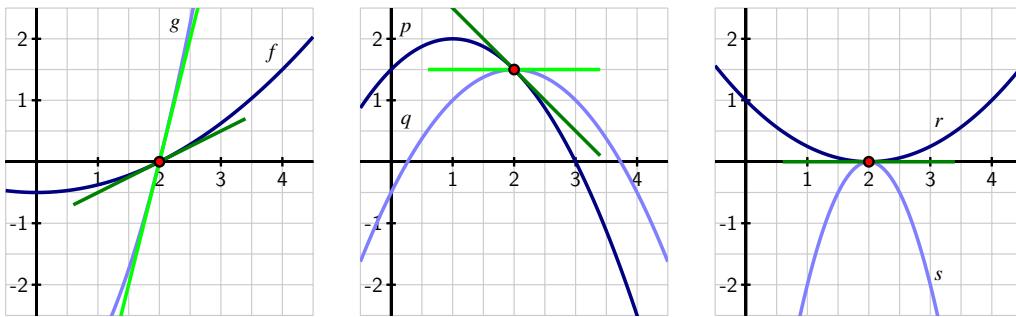


Figure 3.6.6: Three graphs referenced in the questions of Example 3.6.5.

- Use the left-hand graph to determine the values of $f(2)$, $f'(2)$, $g(2)$, and $g'(2)$. Then, evaluate $\lim_{x \rightarrow 2} \frac{f(x)}{g(x)}$.
- Use the middle graph to find $p(2)$, $p'(2)$, $q(2)$, and $q'(2)$. Then, determine the value of $\lim_{x \rightarrow 2} \frac{p(x)}{q(x)}$.
- Assume that r and s are functions whose derivatives for which $r''(2) \neq 0$ and $s''(2) \neq 0$. Use the right-hand graph to compute $r(2)$, $r'(2)$, $s(2)$, $s'(2)$. Explain why you cannot determine the exact value of $\lim_{x \rightarrow 2} \frac{r(x)}{s(x)}$ without further information being provided, but that you can determine the sign of $\lim_{x \rightarrow 2} \frac{r(x)}{s(x)}$. In addition, state what the sign of the limit will be, with justification.

Hint.

- Don't forget that $f'(a)$ measures the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$.
- Do the functions p and q meet the criteria of L'Hopital's Rule?
- Remember that L'Hopital's Rule can be applied more than once to a particular limit.

Answer.

- $\lim_{x \rightarrow 2} \frac{f(x)}{g(x)} = \frac{1}{8}$.

b. $\lim_{x \rightarrow 2} \frac{p(x)}{q(x)} = 1.$

c. $\lim_{x \rightarrow 2} \frac{r(x)}{s(x)} < 0.$

Solution.

- a. From the given graph, we observe that $f(2) = 0$, $f'(2) = \frac{1}{2}$, $g(2) = 0$, and $g'(2) = 4$. By L'Hpital's Rule,

$$\lim_{x \rightarrow 2} \frac{f(x)}{g(x)} = \frac{f'(2)}{g'(2)} = \frac{\frac{1}{2}}{4} = \frac{1}{8}.$$

- b. The given graph tells us that $p(2) = 1.5$, $p'(2) = -1$, $q(2) = 1.5$, and $q'(2) = 0$. Note well that the given limit,

$$\lim_{x \rightarrow 2} \frac{p(x)}{q(x)},$$

is not indeterminate, and thus L'Hpital's Rule does not apply. Rather, since $p(x) \rightarrow 1.5$ and $q(x) \rightarrow 1.5$ as $x \rightarrow 2$, we have that

$$\lim_{x \rightarrow 2} \frac{p(x)}{q(x)} = \frac{p(2)}{q(2)} = \frac{1.5}{1.5} = 1.$$

- c. From the third graph, $r(2) = 0$, $r'(2) = 0$, $s(2) = 0$, $s'(2) = 0$. By L'Hpital's Rule,

$$\lim_{x \rightarrow 2} \frac{r(x)}{s(x)} = \lim_{x \rightarrow 2} \frac{r'(x)}{s'(x)},$$

but this limit is still indeterminate, so by L'Hpital's Rule again,

$$\lim_{x \rightarrow 2} \frac{r(x)}{s(x)} = \lim_{x \rightarrow 2} \frac{r''(x)}{s''(x)} = \frac{r''(2)}{s''(2)},$$

provided that $s''(2) \neq 0$. Since we do not know the values of $r''(2)$ and $s''(2)$, we can't determine the actual value of the limit, but from the graph it appears that $r''(2) > 0$ (since r is concave up) and that $s''(2) < 0$ (because s is concave down), and therefore

$$\lim_{x \rightarrow 2} \frac{r(x)}{s(x)} < 0.$$

3.6.2 Limits involving ∞

The concept of infinity, denoted ∞ , arises naturally in calculus, as 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 3.6.7.

We note that $x = 0$ is not in the domain of f , so we may naturally wonder what happens as $x \rightarrow 0$. As $x \rightarrow 0^+$, 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.

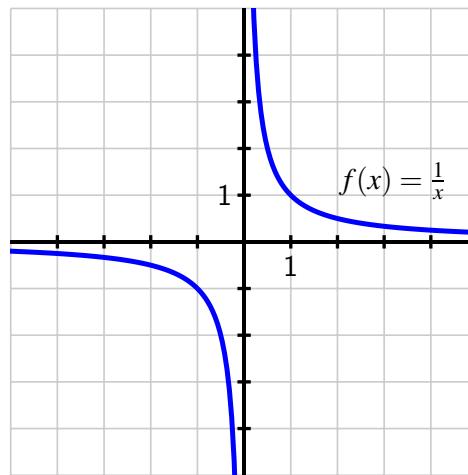


Figure 3.6.7: The graph of $f(x) = \frac{1}{x}$.

Recall that the statement $\lim_{x \rightarrow a} f(x) = L$, 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 now 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

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

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 write

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

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, the notation $\lim_{x \rightarrow a} f(x) = \infty$ means that we can make $f(x)$ as large as we like by taking x sufficiently close (but not equal) to a , and the notation $\lim_{x \rightarrow \infty} f(x) = L$ means that we can make $f(x)$ as close to L as we like by taking x sufficiently large. This notation also applies to left- and right-hand limits, and to limits involving $-\infty$. For example, returning to Figure 3.6.7 and $f(x) = \frac{1}{x}$, we can say that

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

Finally, we write

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

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

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

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

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 $x \rightarrow \infty$ and as $x \rightarrow -\infty$. Here we briefly examine some familiar functions and note the values of several limits involving ∞ .

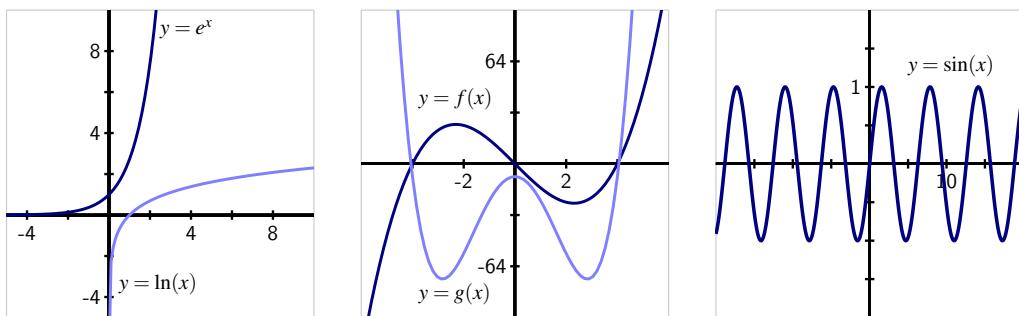


Figure 3.6.8: Graphs of some familiar functions whose end behavior as $x \rightarrow \pm\infty$ is known. In the middle graph, $f(x) = x^3 - 16x$ and $g(x) = x^4 - 16x^2 - 8$.

For the natural exponential function e^x , we note that $\lim_{x \rightarrow \infty} e^x = \infty$ and $\lim_{x \rightarrow -\infty} e^x = 0$. For the exponential decay function e^{-x} , these limits are reversed, with $\lim_{x \rightarrow \infty} e^{-x} = 0$ and $\lim_{x \rightarrow -\infty} e^{-x} = \infty$. Turning to the natural logarithm function, we have $\lim_{x \rightarrow 0^+} \ln(x) = -\infty$ and $\lim_{x \rightarrow \infty} \ln(x) = \infty$. While both e^x and $\ln(x)$ grow without bound as $x \rightarrow \infty$, 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 $\lim_{x \rightarrow \infty} p(x) = \infty$ and $\lim_{x \rightarrow -\infty} p(x) = \infty$, as in the plot of g in Figure 3.6.8. If instead a_n is negative, then $\lim_{x \rightarrow \infty} p(x) = -\infty$ and $\lim_{x \rightarrow -\infty} p(x) = -\infty$. In the situation where n is odd, then either $\lim_{x \rightarrow \infty} p(x) = \infty$ and $\lim_{x \rightarrow -\infty} p(x) = -\infty$ (which occurs when a_n is positive, as in the graph of f in Figure 3.6.8), or $\lim_{x \rightarrow \infty} p(x) = -\infty$ and $\lim_{x \rightarrow -\infty} p(x) = \infty$ (when a_n is negative).

A function can fail to have a limit as $x \rightarrow \infty$. For example, consider the plot of the sine function at right in Figure 3.6.8. Because the function continues oscillating between -1 and 1 as $x \rightarrow \infty$, we say that $\lim_{x \rightarrow \infty} \sin(x)$ does not exist.

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

Example 3.6.9. Determine the limit of the function

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

as $x \rightarrow \infty$.

Note that both $(3x^2 - 4x + 5) \rightarrow \infty$ as $x \rightarrow \infty$ and $(7x^2 + 9x - 10) \rightarrow \infty$ as $x \rightarrow \infty$. Here we say that $\lim_{x \rightarrow \infty} q(x)$ has indeterminate form $\frac{\infty}{\infty}$. 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

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

since $\frac{1}{x^2} \rightarrow 0$ and $\frac{1}{x} \rightarrow 0$ as $x \rightarrow \infty$. 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 $x \rightarrow \infty$.

But how should we handle a limit such as

$$\lim_{x \rightarrow \infty} \frac{x^2}{e^x}?$$

Here, both $x^2 \rightarrow \infty$ and $e^x \rightarrow \infty$, 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 $x \rightarrow a$ (where a is allowed to be ∞), then

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}.$$

(To be technically correct, we need to add the additional hypothesis that $g'(x) \neq 0$ on an open interval that contains a or in every neighborhood of infinity if a is ∞ ; this is almost always met in practice.)

To evaluate $\lim_{x \rightarrow \infty} \frac{x^2}{e^x}$, we can apply L'Hopital's Rule, since both $x^2 \rightarrow \infty$ and $e^x \rightarrow \infty$. Doing so, it follows that

$$\lim_{x \rightarrow \infty} \frac{x^2}{e^x} = \lim_{x \rightarrow \infty} \frac{2x}{e^x}.$$

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, and find that

$$\lim_{x \rightarrow \infty} \frac{x^2}{e^x} = \lim_{x \rightarrow \infty} \frac{2x}{e^x} = \lim_{x \rightarrow \infty} \frac{2}{e^x}.$$

Now, since 2 is constant and $e^x \rightarrow \infty$ as $x \rightarrow \infty$, it follows that $\frac{2}{e^x} \rightarrow 0$ as $x \rightarrow \infty$, which shows that

$$\lim_{x \rightarrow \infty} \frac{x^2}{e^x} = 0.$$

Example 3.6.10. Evaluate each of the following limits. If you use L'Hpital's Rule, indicate where it was used, and be certain its hypotheses are met before you apply it.

a. $\lim_{x \rightarrow \infty} \frac{x}{\ln(x)}$

d. $\lim_{x \rightarrow \frac{\pi}{2}^-} \frac{\tan(x)}{x - \frac{\pi}{2}}$

b. $\lim_{x \rightarrow \infty} \frac{e^x + x}{2e^x + x^2}$

c. $\lim_{x \rightarrow 0^+} \frac{\ln(x)}{\frac{1}{x}}$

e. $\lim_{x \rightarrow \infty} xe^{-x}$

Hint.

- a. Remember that $\ln(x) \rightarrow \infty$ as $x \rightarrow \text{inf}ty$.
- b. Both the numerator and denominator tend to ∞ as $x \rightarrow \infty$.
- c. Note that $x \rightarrow 0^+$, not ∞ .
- d. As $x \rightarrow \frac{\pi}{2}^-$, $\tan(x) \rightarrow \infty$.
- e. Observe that $e^{-x} = \frac{1}{e^x}$.

Answer.

a. $\lim_{x \rightarrow \infty} \frac{x}{\ln(x)} = \infty$.

b. $\lim_{x \rightarrow \infty} \frac{e^x + x}{2e^x + x^2} = \frac{1}{2}$.

c. $\lim_{x \rightarrow 0^+} \frac{\ln(x)}{\frac{1}{x}} = 0$.

d. $\lim_{x \rightarrow \frac{\pi}{2}^-} \frac{\tan(x)}{x - \frac{\pi}{2}} = -\infty$.

e. $\lim_{x \rightarrow \infty} xe^{-x} = 0$.

Solution.

- a. As both numerator and denominator tend to ∞ as $x \rightarrow \infty$, by L'Hpital's Rule followed by some elementary algebra,

$$\lim_{x \rightarrow \infty} \frac{x}{\ln(x)} = \lim_{x \rightarrow \infty} \frac{1}{\frac{1}{x}} = \lim_{x \rightarrow \infty} x = \infty.$$

- b. Because this limit has indeterminate form $\frac{\infty}{\infty}$, L'Hopital's Rule tells us that

$$\lim_{x \rightarrow \infty} \frac{e^x + x}{2e^x + x^2} = \lim_{x \rightarrow \infty} \frac{e^x + 1}{2e^x + 2x}.$$

The latest limit is indeterminate for the same reason, and a second application of the rule shows

$$\lim_{x \rightarrow \infty} \frac{e^x + x}{2e^x + x^2} = \lim_{x \rightarrow \infty} \frac{e^x}{2e^x + 2}.$$

Note how each application of the rule produces a simpler numerator and denominator. With one more use of L'Hopital's Rule, followed by a simple algebraic simplification, we have

$$\lim_{x \rightarrow \infty} \frac{e^x + x}{2e^x + x^2} = \lim_{x \rightarrow \infty} \frac{e^x}{2e^x} = \lim_{x \rightarrow \infty} \frac{1}{2} = \frac{1}{2}.$$

- c. As $x \rightarrow 0^+$, $\ln(x) \rightarrow -\infty$ and $\frac{1}{x} \rightarrow +\infty$, thus by L'Hopital's Rule,

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

Reciprocating, multiplying, and simplifying, it follows that

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

- d. Here, the numerator tends to ∞ while the denominator tends to 0^- . Note well that this limit is not indeterminate, but rather produces a collection of fractions with large positive numerators and small negative denominators. Hence

$$\lim_{x \rightarrow \frac{\pi}{2}^-} \frac{\tan(x)}{x - \frac{\pi}{2}} = -\infty.$$

In particular, we observe that L'Hopital's Rule is not applicable here.

- e. In its original form, $\lim_{x \rightarrow \infty} xe^{-x}$, is indeterminate of form $\infty \cdot 0$. Rewriting e^{-x} as $\frac{1}{e^x}$, a straightforward application of L'Hopital's Rule tells us that

$$\lim_{x \rightarrow \infty} xe^{-x} = \lim_{x \rightarrow \infty} \frac{x}{e^x} = \lim_{x \rightarrow \infty} \frac{1}{e^x}.$$

Since $e^x \rightarrow \infty$ as $x \rightarrow \infty$, we find that

$$\lim_{x \rightarrow \infty} xe^{-x} = 0.$$

To evaluate 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 $x \rightarrow \infty$ provided that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0,$$

whereas f dominates g provided that $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \infty$. Finally, if the value of $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$ is finite and nonzero, we say that f and g *grow at the same rate*. For example, we saw that $\lim_{x \rightarrow \infty} \frac{x^2}{e^x} = 0$, so e^x dominates x^2 , while $\lim_{x \rightarrow \infty} \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10} = \frac{3}{7}$, so $f(x) = 3x^2 - 4x + 5$ and $g(x) = 7x^2 + 9x - 10$ grow at the same rate.

3.6.3 Summary

- Derivatives can be used to help us evaluate indeterminate limits of the form $\frac{0}{0}$ through L'Hpital's Rule, 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'Hpital's Rule tells us that

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}.$$

- When we write $x \rightarrow \infty$, this means that x is increasing without bound. Thus, $\lim_{x \rightarrow \infty} f(x) = L$ means that we can make $f(x)$ as close to L as we like by choosing x to be sufficiently large. Similarly, $\lim_{x \rightarrow a} f(x) = \infty$, means that we can make $f(x)$ as large as we like by choosing x sufficiently close to a .
- A version of L'Hpital's Rule also helps us evaluate indeterminate limits of the form $\frac{\infty}{\infty}$. If f and g are differentiable and both approach zero or both approach $\pm\infty$ as $x \rightarrow a$ (where a is allowed to be ∞), then

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}.$$

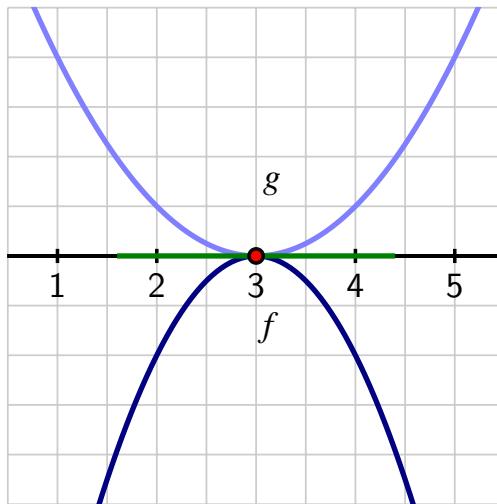
3.6.4 Exercises

- (L'Hpital's Rule with graphs) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (L'Hpital's Rule to evaluate a limit) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Determining if L'Hpital's Rule applies) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Using L'Hpital's Rule multiple times) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- Let f and g be differentiable functions about which the following information is known: $f(3) = g(3) = 0$, $f'(3) = g'(3) = 0$, $f''(3) = -2$, and $g''(3) = 1$. Let a new function h be given by the rule $h(x) = \frac{f(x)}{g(x)}$. On the same set of axes, sketch possible graphs of f and g near $x = 3$, and use the provided information to determine the value of

$$\lim_{x \rightarrow 3} h(x).$$

Provide explanation to support your conclusion.

Answer. $\lim_{x \rightarrow 3} h(x) = -2$.



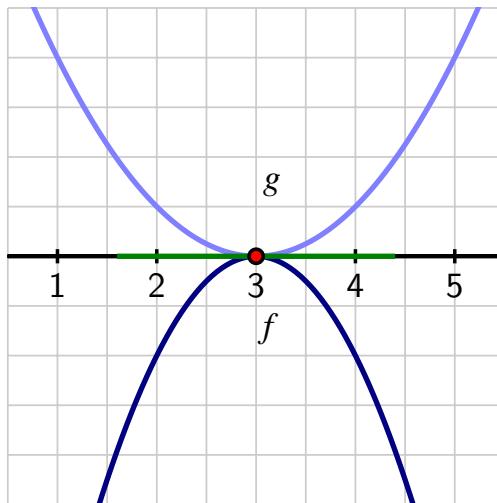
Solution. See the figure below for possible graphs of f and g . To evaluate the limit, we can apply L'Hopital's Rule twice. Since $f(3) = g(3) = 0$, we know by the rule that

$$\lim_{x \rightarrow 3} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 3} \frac{f'(x)}{g'(x)}.$$

Since $f'(3) = g'(3) = 0$, the limit is still indeterminate, and we can apply the rule again to get

$$\lim_{x \rightarrow 3} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 3} \frac{f''(x)}{g''(x)} = \frac{f''(3)}{g''(3)} = \frac{-2}{1}.$$

Hence, $\lim_{x \rightarrow 3} h(x) = -2$.

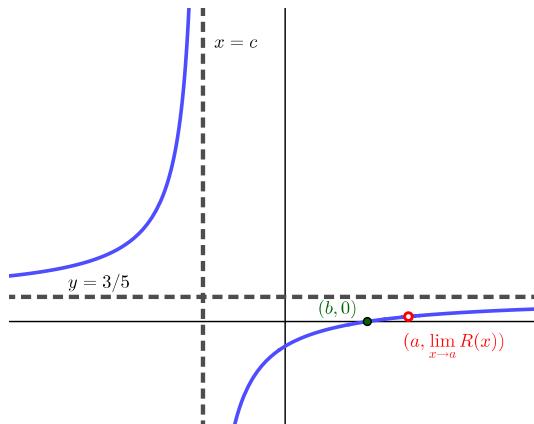


6. Find all vertical and horizontal asymptotes of the function

$$R(x) = \frac{3(x-a)(x-b)}{5(x-a)(x-c)},$$

where a , b , and c are distinct, arbitrary constants. In addition, state all values of x for which R is not continuous. Sketch a possible graph of R , clearly labeling the values of a , b , and c .

Answer. Horizontal asymptote: $y = \frac{3}{5}$; vertical asymptote: $x = c$; hole: $(a, \frac{3(a-b)}{5(a-c)})$. R is not continuous at $x = a$ and $x = c$.



Solution. First, we note that since both the numerator and denominator of R are quadratic, the limit as $x \rightarrow \infty$ can be found by taking the quotient of the leading coefficients. In particular,

$$\lim_{x \rightarrow \infty} R(x) = \frac{3}{5}$$

and thus the horizontal asymptote of R is $y = \frac{3}{5}$.

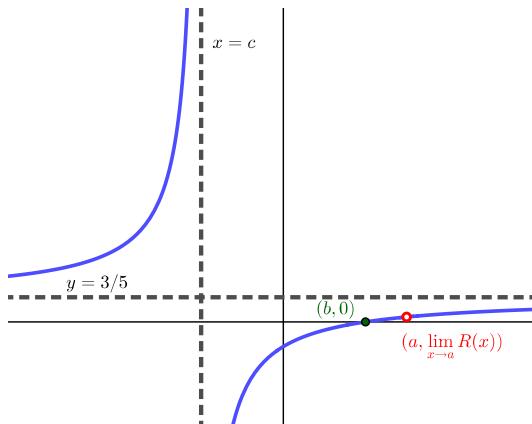
Next, since b and c are distinct, we see that the numerator of R is zero (and the denominator is not) when $x = b$, and thus $R(b) = 0$. In addition, the denominator is zero (and the numerator is not) when $x = c$, and this results in there being a vertical asymptote at $x = c$.

Finally, we observe that $x = a$ makes both the numerator and denominator zero. Since $R(a)$ is thus not defined, the graph of R has either a hole or a vertical asymptote at $x = a$. To determine which, we evaluate the limit and see that

$$\lim_{x \rightarrow a} \frac{3(x-a)(x-b)}{5(x-a)(x-c)} = \lim_{x \rightarrow a} \frac{3(x-b)}{5(x-c)} = \frac{3(a-b)}{5(a-c)}.$$

This tells us that R has a hole at the point $(a, \frac{3(a-b)}{5(a-c)})$. Moreover, R is not continuous at $x = a$ and $x = c$.

A possible graph of R is shown in the following figure.



7. Consider the function $g(x) = x^{2x}$, which is defined for all $x > 0$. Observe that $\lim_{x \rightarrow 0^+} g(x)$ is indeterminate due to its form of 0^0 . (Think about how we know that $0^k = 0$ for all $k > 0$, while $b^0 = 1$ for all $b \neq 0$, but that neither rule can apply to 0^0 .)

- Let $h(x) = \ln(g(x))$. Explain why $h(x) = 2x \ln(x)$.
- Next, explain why it is equivalent to write $h(x) = \frac{2 \ln(x)}{\frac{1}{x}}$.
- Use L'Hopital's Rule and your work in (b) to compute $\lim_{x \rightarrow 0^+} h(x)$.
- Based on the value of $\lim_{x \rightarrow 0^+} h(x)$, determine $\lim_{x \rightarrow 0^+} g(x)$.

Answer.

- $\ln(x^{2x}) = 2x \cdot \ln(x)$.
- $x = \frac{1}{\frac{1}{x}}$.
- $\lim_{x \rightarrow 0^+} h(x) = 0$.
- $\lim_{x \rightarrow 0^+} g(x) = \lim_{x \rightarrow 0^+} x^{2x} = 1$.

Solution.

- Since $h(x) = \ln(g(x))$ and we are given that $g(x) = x^{2x}$, it follows that

$$h(x) = \ln(x^{2x}) = 2x \cdot \ln(x).$$

- Since $x = (x^{-1})^{-1} = \frac{1}{\frac{1}{x}}$, we can also write

$$h(x) = \frac{2 \ln(x)}{\frac{1}{x}}.$$

- Since $\ln(x) \rightarrow -\infty$ as $x \rightarrow 0^+$ and $\frac{1}{x} \rightarrow +\infty$ as $x \rightarrow 0^+$, we see that

$$\lim_{x \rightarrow 0^+} h(x) = \lim_{x \rightarrow 0^+} \frac{2 \ln(x)}{\frac{1}{x}}$$

is of an indeterminate form to which we can apply L'Hopital's Rule. Doing so,

$$\lim_{x \rightarrow 0^+} h(x) = \lim_{x \rightarrow 0^+} \frac{2 \ln(x)}{\frac{1}{x}} = \lim_{x \rightarrow 0^+} \frac{\frac{2}{x}}{-\frac{1}{x^2}}$$

Simplifying algebraically,

$$\lim_{x \rightarrow 0^+} h(x) = \lim_{x \rightarrow 0^+} \frac{2}{x} \cdot \left(-\frac{x^2}{1} \right) = \lim_{x \rightarrow 0^+} (-2x)$$

Hence, it follows that $\lim_{x \rightarrow 0^+} h(x) = 0$.

- d. Having shown that $\lim_{x \rightarrow 0^+} h(x) = 0$ and recalling that $h(x) = \ln(g(x))$, we know that as $x \rightarrow 0^+$, the natural log of $g(x)$ tends to 0. Since $\ln(1) = 0$, it must be the case that $g(x)$ is tending to 1, and thus

$$\lim_{x \rightarrow 0^+} g(x) = \lim_{x \rightarrow 0^+} x^{2x} = 1$$

8. Recall we say that function g **dominates** function f provided that $\lim_{x \rightarrow \infty} f(x) = \infty$, $\lim_{x \rightarrow \infty} g(x) = \infty$, and $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$.

- Which function dominates the other: $\ln(x)$ or \sqrt{x} ?
- Which function dominates the other: $\ln(x)$ or $\sqrt[n]{x}$? (n can be any positive integer)
- Explain why e^x will dominate any polynomial function.
- Explain why x^n will dominate $\ln(x)$ for any positive integer n .
- Give any example of two nonlinear functions such that neither dominates the other.

Answer.

- Show that $\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt{x}} = 0$.
- Show that $\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt[n]{x}} = 0$.
- Consider $\lim_{x \rightarrow \infty} \frac{p(x)}{e^x}$. By repeated application of LHR, the numerator will eventually be simply a constant (after n applications of LHR), and thus with e^x still in the denominator, the overall limit will be 0.
- Show that $\lim_{x \rightarrow \infty} \frac{\ln(x)}{x^n} = 0$
- For example, $f(x) = 3x^2 + 1$ and $g(x) = -0.5x^2 + 5x - 2$.

Solution.

- a. We note that both $\ln(x)$ or \sqrt{x} increase without bound as $x \rightarrow \infty$, so we consider

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt{x}}.$$

By L'Hopital's Rule and some simplifying algebra,

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt{x}} = \lim_{x \rightarrow \infty} \frac{\frac{1}{x}}{\frac{1}{2x^{1/2}}} = \lim_{x \rightarrow \infty} \frac{2x^{1/2}}{x} = \lim_{x \rightarrow \infty} \frac{2}{x^{1/2}}.$$

Thus, it follows that

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt{x}} = 0,$$

and hence \sqrt{x} dominates $\ln(x)$.

- b. Using a similar argument to our work in (a), we can show that $\sqrt[n]{x}$ dominates $\ln(x)$ for any positive integer n . In particular, by L'Hopital's Rule and some simplifying algebra,

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt[n]{x}} = \lim_{x \rightarrow \infty} \frac{\frac{1}{x}}{\frac{1}{nx^{(n-1)/n}}} = \lim_{x \rightarrow \infty} \frac{nx^{(n-1)/n}}{x}.$$

Since $\frac{x^{(n-1)/n}}{x} = x^{(n-1)/n-1} = x^{(n-1-n)/n} = x^{-1/n}$, it follows that

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{\sqrt[n]{x}} = \lim_{x \rightarrow \infty} \frac{1}{x^{1/n}} = 0$$

we've shown that $\sqrt[n]{x}$ dominates $\ln(x)$ for any positive integer n .

- c. To see why e^x will dominate any polynomial function, consider $p(x) = x^4$. If we consider the limit of their quotient and apply LHR once, we see that

$$\lim_{x \rightarrow \infty} \frac{x^4}{e^x} = \lim_{x \rightarrow \infty} \frac{4x^3}{e^x}$$

By repeated application of LHR, the numerator (regardless of what polynomial we start with) will eventually be simply a constant (after n applications of LHR), and thus with e^x still in the denominator, the overall limit will be 0.

- d. Here we naturally consider the limit of the quotient of $\ln(x)$ and x^n , where $n \geq 1$. Since both functions increase without bound, we can apply LHR. Doing so along with some simplifying algebra, we see

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{x^n} = \lim_{x \rightarrow \infty} \frac{\frac{1}{x}}{nx^{n-1}} = \frac{1}{nx^n} = 0$$

and thus x^n indeed dominates.

- e. Consider $f(x) = 3x^2 + 1$ and $g(x) = -0.5x^2 + 5x - 2$. Since these functions are polynomials of the same degree, we know

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \frac{3}{-0.5} = -6.$$

3.7 Parametric Equations

Motivating Questions

- What types of equations can be used to describe motion in multiple dimensions as a function of time?
 - How can we calculate instantaneous rates of change for such equations?
 - How can we derive the equations of lines tangent to such equations?
-

Up to this point, we have been representing a graph by a single equation involving two variables such as x and y . However, some items are not really well described by such equations. In this section, you will study situations in which it is useful to introduce a *third variable* to represent a curve in the plane. This process is commonly called parameterization and is the basis for our study of parametric curves.

3.7.1 Parametric Equations

Consider the path of an object that is propelled into the air at an angle of 45° . In a physics class you might learn that this object would follow a parabolic path. That is

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

However, this equation does not tell the whole story. Although it does tell you where the object has been, it does not tell you *when* the object was at a given point (x, y) on the path. To determine this time, you can introduce a third variable t , called a parameter. It is possible to write both x and y as functions of t to obtain the parametric equations

$$\begin{aligned}x(t) &= 24\sqrt{2}t \\y(t) &= -16t^2 + 24\sqrt{2}t\end{aligned}$$

The parametric equations are graphed in Figure 3.7.1 below. Using the parametric equations we can state properties such as: at time $t = 0$, the object is at the point $(0, 0)$ and at time $t = 1$, the object is at the point $(24\sqrt{2}, 24\sqrt{2} - 16)$.

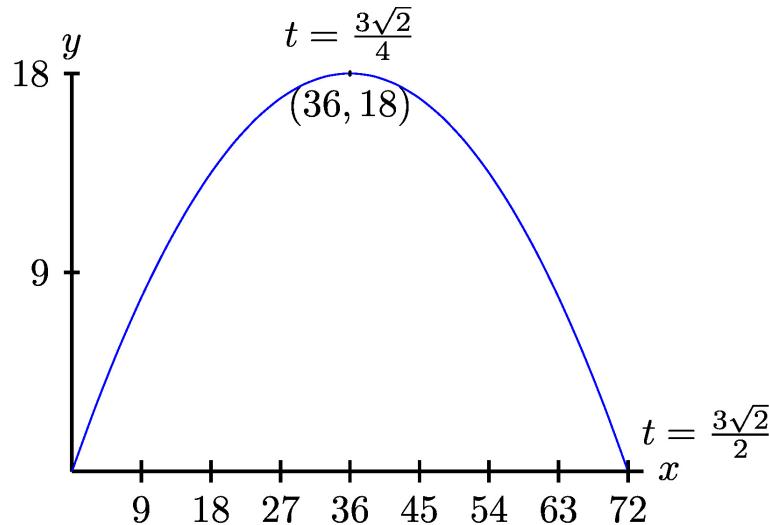


Figure 3.7.1: A plot of the parametric equations $x(t) = 24\sqrt{2}t$, $y(t) = -16t^2 + 24\sqrt{2}t$. Notice that when $t = \frac{3\sqrt{2}}{4}$, $x = 36$ and $y = 18$ (you can get these values by plugging $t = \frac{3\sqrt{2}}{4}$ into the equations for x and y).

Example 3.7.2. Suppose the position of an object is given by the parametric equations

$$\begin{aligned}x(t) &= 24\sqrt{2}t \\y(t) &= -16t^2 + 24\sqrt{2}t.\end{aligned}$$

After $t = 3$ seconds, what is the position of the object? If the origin represents the initial position?

Hint. Try plugging $t = 3$ into each equation.

Solution. By plugging $t = 3$ into each of the equations for $x(t)$ and $y(t)$ we see that

$$\begin{aligned}x(3) &= 24\sqrt{2}3 \\y(3) &= -16(3)^2 + 24\sqrt{2}(3)\end{aligned}$$

Therefore, after 3 seconds the object is located at $(24\sqrt{2}3, 144 - 72\sqrt{2})$.

With these examples in hand we are now in a position to formally define parametric equations.

Parametric Equations

Parametric equations are equations that specify the values of x and y in terms of a third variable t called a *parameter*. We often represent parametric curves in the form

$$x(t) = f(t)$$

$$y(t) = g(t).$$

where f and g are functions and the parameter t varies over some interval $a < t < b$. When no interval is specified we assume that t accepts all real values.

3.7.2 Graphing Parametric Equations

Graphs of curves sketched from parametric equations can have varied and interesting shapes, as exemplified in Figure 3.7.3. In this section we will cover some methods to sketch parametric curves.

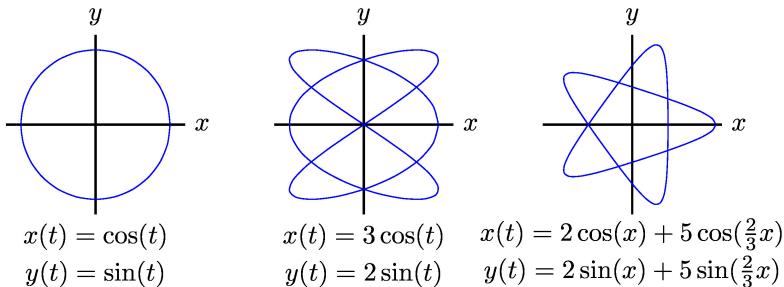


Figure 3.7.3: Parametric equations can give some very interesting graphs.

One of the most effective ways to sketch a parametric curve is to create a table of values by choosing various values of t and computing both $x(t)$ and $y(t)$. The following example demonstrates this procedure.

Example 3.7.4. Sketch the curve given by the parametric equations

$$x(t) = t^2 - 4$$

$$y(t) = t/2.$$

Hint. A reasonable first step is to create a table of values for $t = -2, -1, 0, 1, 2, 3$. Why these values of t ? Choosing "good" values of t is not always a straight forward task and in general you must choose values that give you a "reasonable" picture of the curve. You also might wonder how many values of t are appropriate. Again, the answer is enough values that you can be reasonably confident you have captured the shape of the curve. So what does this mean? Often it helps to choose a few positive and negative values close to zero, calculate these values, plot the points, and then determine if you need more values to give your curve a clear shape.

Solution. We begin by creating a table of values (For more details on how we chose these values see the hint).

t	x(t)	y(t)
-2	0	-1
-1	-3	-1/2
0	-4	0
1	-3	1/2
2	0	1
3	5	3/2

Table 3.7.5: Table of values for the parametric equations $x(t) = t^2 - 4$ and $y(t) = t/2$

We plot these values on a coordinate plane and then connect the values with appropriate curves as in Figure 3.7.6. It is often helpful to place arrows on the curve to indicate how the curve changes as t increases.

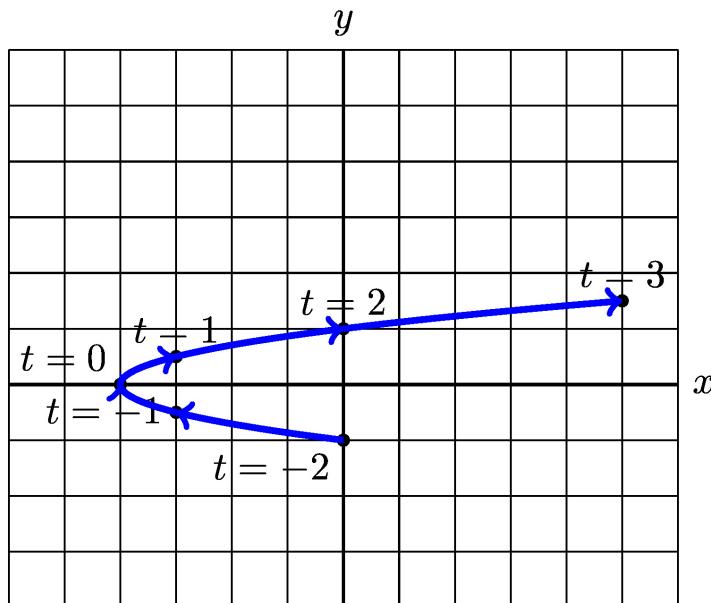


Figure 3.7.6: Plot of the parametric equations $x(t) = t^2 - 4$ and $y(t) = t/2$ for $-2 < t < 3$.

You may notice in Figure 3.7.6 that the curve seems to start abruptly at $t = -2$. In fact, this is just an artifact of our starting our table of values at $t = -2$. If we instead started our table at $t = -3$ then we would get a slightly different image, Figure 3.7.7.

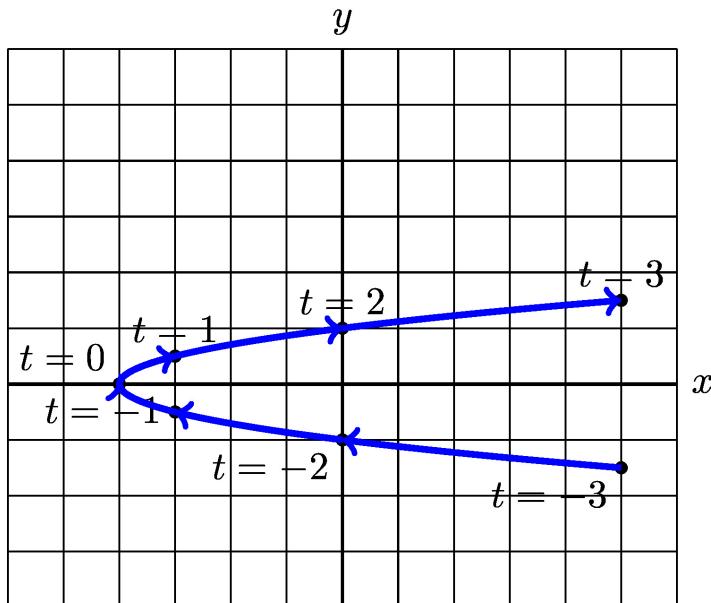


Figure 3.7.7: Plot of the parametric equations $x(t) = t^2 - 4$ and $y(t) = t/2$ for $-3 < t < 3$.

In addition to the method of plotting points, many graphing software programs can effectively plot parametric curves. For convenience, one such utility based on GeoGebra is included in Figure 3.7.8.

Geogebra: www.geogebra.org/m/pcj3yebb

Figure 3.7.8: GeoGebra app for plotting parametric equations.

Example 3.7.9. Describe the motion of the particle whose coordinates at time t are expressed with the parametric equations: $x = \cos t$ and $y = \sin t$.

Solution. Since $(\cos t)^2 + (\sin t)^2 = 1$, we have $x^2 + y^2 = 1$. That is, at any time t the particle is at a point (x, y) on the unit circle $x^2 + y^2 = 1$. We plot points at different times to see how the particle moves on the circle. The particle completes one full trip counterclockwise around the circle every 2π units of time. Notice how the x -coordinate goes back and forth from -1 to 1 while the y -coordinate goes up and down from -1 to 1 . The two motions combine to trace out a circle.

t	x	y
0	1	0
$\pi/2$	0	1
π	-1	0
$3\pi/2$	0	-1
2π	1	0

Table 3.7.10: Table of values for the parametric equations $x = \cos t$ and $y = \sin t$.

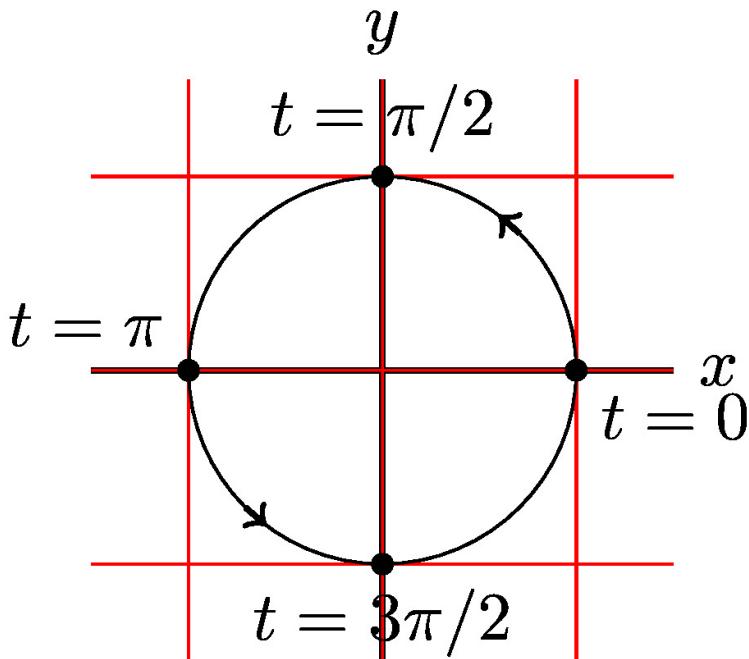


Figure 3.7.11: Plot of the parametric equations $x = \cos t$ and $y = \sin t$.

The first thing we note is that $(\cos 3t)^2 + (\sin 3t)^2 = 1$ and so we still have the fact that, at any time t the particle is at a point (x, y) on the unit circle $x^2 + y^2 = 1$. So, what effect is the 3 in $x = \cos 3t$ and $y = \sin 3t$ having? Consider the following table of values:

t	x	y
0	1	0
$\pi/6$	0	1
$\pi/3$	-1	0
$\pi/2$	0	-1
$2\pi/3$	1	0

Table 3.7.12: Table of values for the parametric equations $x = \cos 3t$ and $y = \sin 3t$.

In other words, the motion of the particle described by $x = \cos 3t$ and $y = \sin 3t$ is nearly identical to the motion of a particle described by $x = \cos t$ and $y = \sin t$ except that the particle described by $x = \cos 3t$ and $y = \sin 3t$ is moving three times as fast. The motion of the particle is displayed by the graph in Figure 3.7.13.

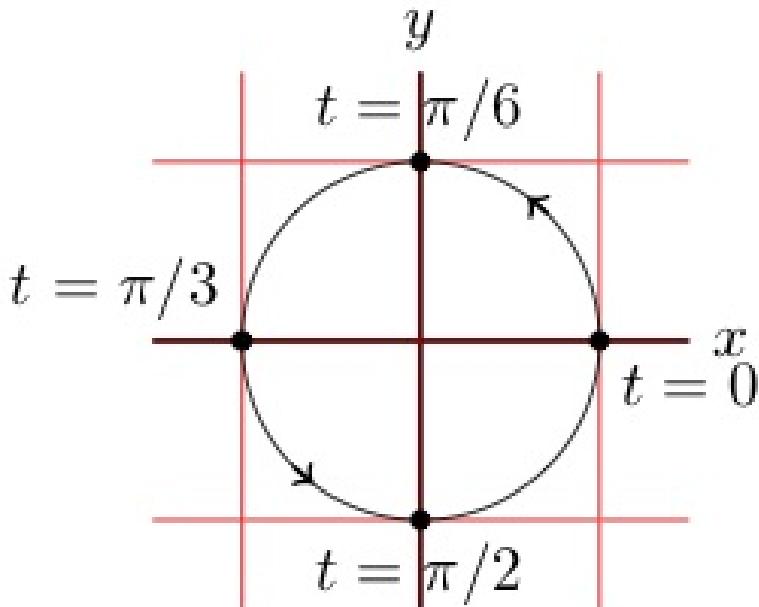


Figure 3.7.13: Plot of the parametric equations $x = \cos 3t$ and $y = \sin 3t$.

3.7.3 Motion in a Straight Line and Derivatives

Suppose an object moves with constant speed along a straight line through the point (x_0, y_0) . Recall that the key to motion in a straight line is that the rate of change is constant. In other words, both the x and y -coordinates have a constant rate of change. For convenience, let $a = dx/dt$ and $b = dy/dt$ be the rate of change in the x and y directions, respectively. Then at time t the object has coordinates $x = x_0 + at$, $y = y_0 + bt$. Notice that a represents the change in x in one unit of time, and b represents the change in y in one unit of time (see Figure 3.7.14). Thus the line has slope $m = b/a$. Therefore, the parametric equations for motion in a straight line are given by $x = x_0 + at$, $y = y_0 + bt$.

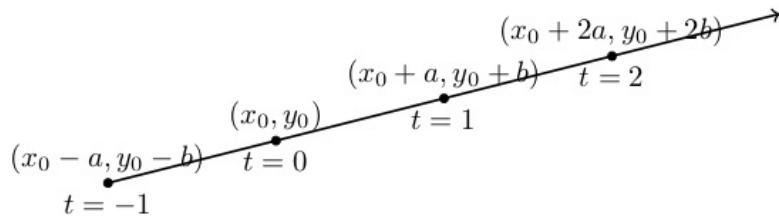


Figure 3.7.14: Plot of the parametric equations $x = \cos 3t$ and $y = \sin 3t$.

Now that we have developed the notion of motion in a straight line for parametric equations we can discuss tangent lines for parametric equations.

Parametric Equations for the tangent line

An object moving along a line through the point (x_0, y_0) , with $dx/dt = a$ and $dy/dt = b$, has parametric equations

$$x = x_0 + at, \quad y = y_0 + bt.$$

The slope of the line is $m = b/a$.

Example 3.7.15. Find the tangent line at the point $(1, 2)$ to the curve defined by the parametric equations $x = t^3$, $y = 2t$.

Solution. At time $t = 1$ the particle is at the point $(1, 2)$. The velocity in the x -direction at time t is $v_x = dx/dt = 3t^2$, and the velocity in the y -direction is $v_y = dy/dt = 2$. So at $t = 1$ the velocity in the x -direction is 3 and the velocity in the y -direction is 2. Thus the tangent line has parametric equations

$$x = 1 + 3t, \quad y = 2 + 2t.$$

With the notion of a tangent line in hand we are in the position to be able to talk about the speed and velocity of an object whose position is given by parametric equations. In general, given an object traveling in the xy -plane we can break its movement into two pieces. Movement in the “ x -direction” and movement in the “ y -direction.” We can also talk about rate-of-change in the “ x -direction” ($\frac{dx}{dt}$) and rate-of-change in the “ y -direction” ($\frac{dy}{dt}$). If we think of these rates of change as vectors then we can use the Pythagorean theorem to talk about the instantaneous speed $\frac{dy}{dx}$, see Figure 3.7.16.

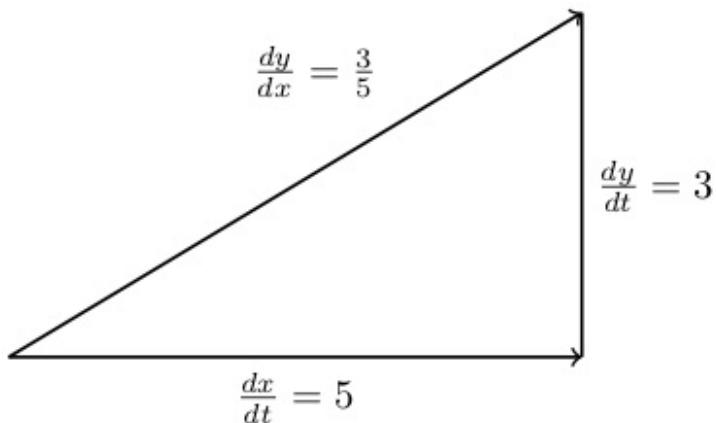


Figure 3.7.16: Plot of the parametric equations $x = \cos 3t$ and $y = \sin 3t$.

Using this idea we have the following definition.

Instantaneous Speed

The instantaneous speed of a moving object is defined to be

$$v = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$$

The quantity $v_x = dx/dt$ is the instantaneous velocity in the x -direction; $v_y = dy/dt$ is the instantaneous velocity in the y -direction. The velocity vector \mathbf{v} is written

$$\mathbf{v} = v_x \mathbf{i} + v_y \mathbf{j}.$$

Example 3.7.17. A particle moves along a curve in the xy -plane with $x(t) = 2t + e^t$ and $y(t) = 3t - 4$, where t is time. Find the velocity vector and speed of the particle when $t = 1$.

Solution. Differentiating gives $dx/dt = 2 + e^t$, $dy/dt = 3$. When $t = 1$ we have $v_x = 2 + e$, $v_y = 3$. So the velocity vector is

$$\mathbf{v} = (2 + e) \mathbf{i} + 3 \mathbf{j}$$

The speed at $t = 1$ is

$$v = \sqrt{(2 + e)^2 + 3^2} \approx 5.591.$$

We now formally define the slope of the tangent line to a parametric curve and concavity of a parametric curve.

Slope of Tangent Line to Parametric Curves

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$

Concavity of Parametric Curves

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt} \left(\frac{dy}{dx} \right)}{\frac{dx}{dt}}$$

Example 3.7.18. A particle moves in the xy -plane so that its position at time t is given by $x = \sin(2t)$, $y = -\cos(2t)$ for $0 \leq t < 2\pi$.

1. At what time does the particle first touch the x -axis? What is the speed of the particle at that time?
2. Is the particle ever at rest?

Solution.

1. The particle touches the x -axis when $y = 0$, that is when $\cos(2t) = 0$. This occurs when $t = \pi/4$. At this point, the speed of the particle is given by

$$\sqrt{4\cos^2(\pi/2) + 4\sin^2(\pi/2)} = 2.$$

2. Since $\sqrt{4\cos^2(2t) + 4\sin^2(2t)} = 2$, the particle is never at rest.

3.7.4 Exercises

1. (Evaluating Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Sketching Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Parametric Lines) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Slope in Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Derivatives of Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Tangent Lines and Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Horizontal Tangent Lines and Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. (Speed and Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

9. (Calculating Speed Using Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

10. (Intersecting Parametric Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

Chapter 4

The Definite Integral

4.1 Determining distance traveled from velocity

Motivating 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?
 - What does it mean to antiderivative a function and why is this process relevant to finding distance traveled?
 - If velocity is negative, how does this impact the problem of finding distance traveled?
-

In the first section of the text, we considered a moving object with known position at time t , namely, a tennis ball tossed into the air with height s (in feet) at time t (in seconds) given by $s(t) = 64 - 16(t - 1)^2$. We investigated the average velocity of the ball on an interval $[a, b]$, computed by the difference quotient $\frac{s(b) - s(a)}{b - a}$. We found that we could determine the instantaneous velocity of the ball at time t by taking the derivative of the position function,

$$s'(t) = \lim_{h \rightarrow 0} \frac{s(t + h) - s(t)}{h}.$$

Thus, if its position function is differentiable, we can find the velocity of a moving object at any point in time.

From this study of position and velocity we have learned a great deal. We can use the derivative to find a function's instantaneous rate of change at any point in the domain, to find

where the function is increasing or decreasing, where it is concave up or concave down, and to locate 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. For all these tasks, we proceed from a function f to its derivative, f' , and use the meaning of the derivative to help us answer important questions.

We have also encountered the reverse situation, where we know the derivative of a function, f' , and try to deduce information about f . We will focus our attention in Chapter 4 on this problem: if we know the instantaneous rate of change of a function, can we find the function itself? We start with a more specific question: if we know the instantaneous velocity of an object moving along a straight line path, can we find its corresponding position function?

Example 4.1.1. 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 left-hand axes provided in Figure 4.1.2, sketch a labeled graph of the velocity function $v(t) = 3$.

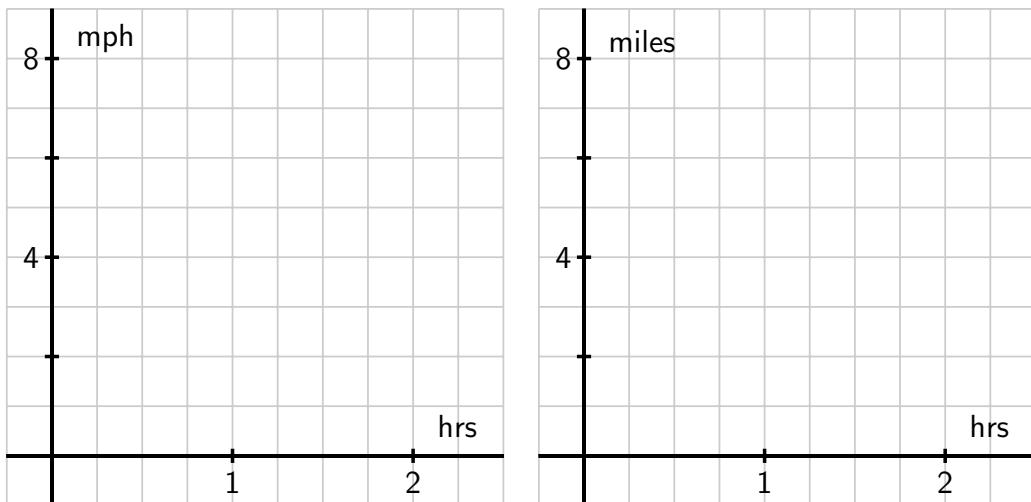


Figure 4.1.2: At left, axes for plotting $y = v(t)$; at right, for plotting $y = s(t)$.

Note that while the scale on the two sets of axes is the same, the units on the right-hand axes differ from those on the left. The right-hand axes will be used in question (d).

- 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 right-hand axes provided in Figure 4.1.2, 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 .

4.1.1 Area under the graph of the velocity function

In Example 4.1.1, we learned that when the velocity of a moving object's velocity is constant (and positive), the area under the velocity curve over an interval of time tells us the distance the object traveled.

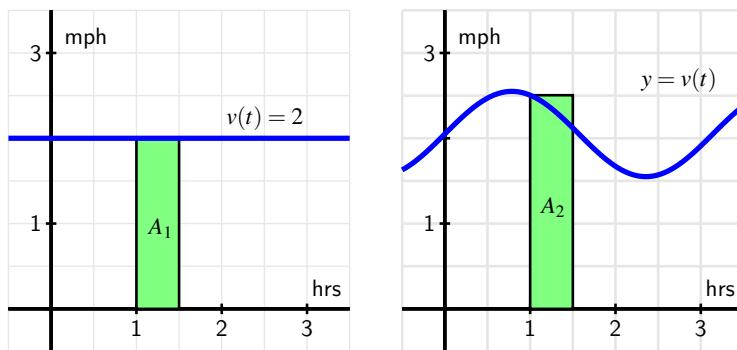


Figure 4.1.3: At left, a constant velocity function; at right, a non-constant velocity function.

The left-hand graph of Figure 4.1.3 shows the velocity of an object moving at 2 miles per hour over the time interval $[1, 1.5]$. 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 result is simply the fact that distance equals rate times time, provided the rate is constant. Thus, if $v(t)$ is constant on the interval $[a, b]$, the distance traveled on $[a, b]$ is equal to the area A given by

$$A = v(a)(b - a) = v(a)\Delta t,$$

where Δt is the change in t over the interval. (Since the velocity is constant, we can use any value of $v(t)$ on the interval $[a, b]$, 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>.¹

The situation is more complicated when the velocity function is not constant. But on relatively small intervals where $v(t)$ does not vary much, we can use the area principle to estimate the distance traveled. The graph at right in Figure 4.1.3 shows a non-constant velocity function. On the interval $[1, 1.5]$, the velocity varies from $v(1) = 2.5$ down to $v(1.5) \approx 2.1$. One estimate for the distance traveled is the area of the pictured rectangle,

$$A_2 = v(1)\Delta t = 2.5 \frac{\text{miles}}{\text{hour}} \cdot \frac{1}{2} \text{ hours} = 1.25 \text{ miles.}$$

¹Marc Renault, calculus applets.

Note that because v is decreasing on $[1, 1.5]$, $A_2 = 1.25$ is an over-estimate of the actual distance traveled.

To estimate the area under this non-constant velocity function on a wider interval, say $[0, 3]$, one rectangle will not give a good approximation. Instead, we could use the six rectangles pictured in Figure 4.1.4, 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 the velocity is sometimes negative? Can we find the exact area under any non-constant curve?

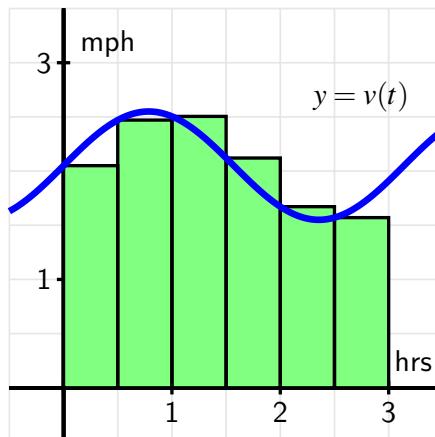


Figure 4.1.4: Using six rectangles to estimate the area under $y = v(t)$ on $[0, 3]$.

We will study these questions and more in what follows; for now it suffices to observe that the simple idea of the area of a rectangle gives us a powerful tool for estimating distance traveled from a velocity function, as well as for estimating the area under an arbitrary curve. To explore the use of multiple rectangles to approximate area under a non-constant velocity function, see the applet found at <http://gvsu.edu/s/9U>.²

Example 4.1.5. Suppose that a person is walking in such a way that her velocity varies slightly according to the information given in Table 4.1.6 and graph given in Figure 4.1.7.

²Marc Renault, calculus applets.

t	$v(t)$
0.00	1.500
0.25	1.789
0.50	1.938
0.75	1.992
1.00	2.000
1.25	2.008
1.50	2.063
1.75	2.211
2.00	2.500

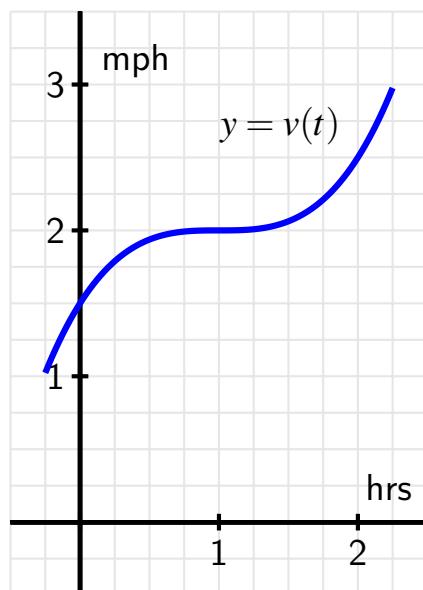


Table 4.1.6: Velocity data for the person walking.

Figure 4.1.7: The graph of $y = v(t)$.

- 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 $\Delta t = 0.5$, choosing a way to use the function consistently to determine the height of each rectangle in order to approximate distance traveled.
- How could you get a better approximation of the distance traveled on $[0, 2]$? Explain, and then find this new estimate.
- 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 $s' = v$.
- Based on your work in (c), what is the value of $s(2) - s(0)$? What is the meaning of this quantity?

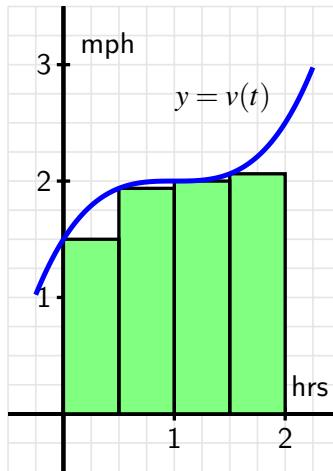
Hint.

- For instance, the approximate distance traveled on $[0, 0.5]$ can be computed by $v(0) \cdot 0.5 = 1.5 \cdot 0.5 = 0.75$ miles.
- Think about the possibility of using a larger number of rectangles.
- If $v(t) = t^3$ and we seek a function s such that $s' = v$, note that s has to involve t^4 .

- d. Observe that this quantity is measuring a change in position.

Answer.

a.



$$\begin{aligned}
 A &= v(0.0) \cdot 0.5 + v(0.5) \cdot 0.5 + v(1.0) \cdot 0.5 + v(1.5) \cdot 0.5 \\
 &= 1.500 \cdot 0.5 + 1.9375 \cdot 0.5 + 2.000 \cdot 0.5 + 2.0625 \cdot 0.5 \\
 &= 3.75
 \end{aligned}$$

Thus, $D \approx 3.75$ miles.

- b. Using 8 rectangles of width 0.25, $D \approx 3.875$.

c. $s(t) = \frac{1}{8}t^4 - \frac{1}{2}t^3 + \frac{3}{4}t^2 + \frac{3}{2}t$.

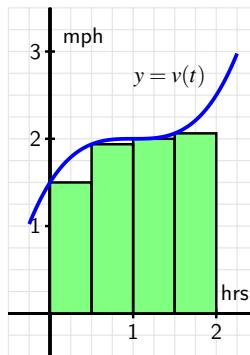
d. $s(2) - s(0) = \frac{1}{8}2^4 - \frac{1}{2}2^3 + \frac{3}{4}2^2 + \frac{3}{2}2 = 4$.

Solution.

- a. Using rectangles of width $\Delta t = 0.5$ and choosing to set the heights of the rectangles from the function value at the left end of the interval, we see the following graph and find the sum of the areas of the rectangles to be

$$\begin{aligned}
 A &= v(0.0) \cdot 0.5 + v(0.5) \cdot 0.5 + v(1.0) \cdot 0.5 + v(1.5) \cdot 0.5 \\
 &= 1.500 \cdot 0.5 + 1.9375 \cdot 0.5 + 2.000 \cdot 0.5 + 2.0625 \cdot 0.5 \\
 &= 3.75
 \end{aligned}$$

Thus, the distance traveled is approximately $D \approx 3.75$ miles.



- b. It appears that a better approximation could be found using narrower rectangles. If we move to 8 rectangles of width 0.25, similar computations show that $D \approx 3.875$.
- c. By thinking about how the power rule for differentiation works, we can undo this rule and find a position function s whose derivative is v . For instance, since $\frac{d}{dt}[t^4] = 4t^3$, we see that $\frac{d}{dt}[\frac{1}{8}t^4] = \frac{1}{2}t^3$. Thus, if we let

$$s(t) = \frac{1}{8}t^4 - \frac{1}{2}t^3 + \frac{3}{4}t^2 + \frac{3}{2}t,$$

then it is straightforward to check that $s'(t) = \frac{1}{2}t^3 - \frac{3}{2}t^2 + \frac{3}{2}t + \frac{3}{2}$, which is precisely the formula for $v(t)$ that we were given.

- d. By the rule found in (c) for s , we have that $s(2) - s(0) = \frac{1}{8}2^4 - \frac{1}{2}2^3 + \frac{3}{4}2^2 + \frac{3}{2}2 = 2 - 4 + 3 + 3 = 4$. This is the change in the walker's position over the time interval $[0, 2]$, and since the velocity is always positive, this is actually the exact distance traveled. We see how both earlier estimates (3.75 and 3.785) are good approximations to this value.

4.1.2 Two approaches: area and antiderivatives

When the velocity of a moving object is positive, the object's position is always increasing. (We will soon consider situations where velocity is negative; for now, we focus on the situation where velocity is always positive.) We have established that whenever v is constant on an interval, the exact distance traveled is the area under the velocity curve. When v is not constant, we can estimate the total distance traveled by finding the areas of rectangles that approximate the area under the velocity curve.

Thus, we see that finding the area between a curve and the horizontal axis is an important exercise: besides being an interesting geometric question, if the curve gives the velocity of a moving object, the area under the curve tells us the exact distance traveled on an interval. We can estimate this area if we have a graph or a table of values for the velocity function.

In Example 4.1.5, we encountered an alternate approach to finding the distance traveled. If $y = v(t)$ is a formula for the instantaneous velocity of a moving object, then v must be the derivative of the object's position function, s . If we can find a formula for $s(t)$ from the

formula for $v(t)$, we will know the position of the object at time t , and the change in position over a particular time interval tells us the distance traveled on that interval.

For a simple example, consider the situation from Example 4.1.1, where a person is walking along a straight line with velocity function $v(t) = 3$ mph.

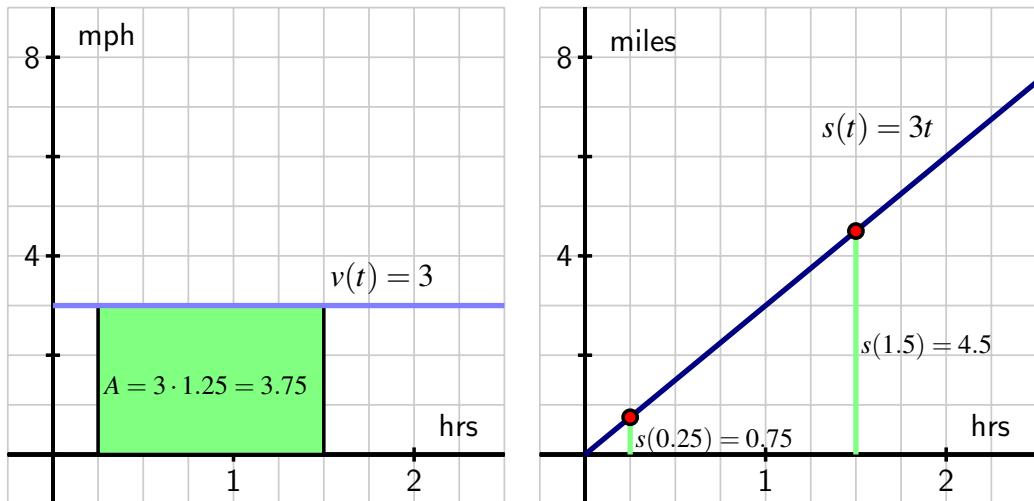


Figure 4.1.8: The velocity function $v(t) = 3$ and corresponding position function $s(t) = 3t$.

On the left-hand graph of the velocity function in Figure 4.1.8, we see the relationship between area and distance traveled,

$$A = 3 \frac{\text{miles}}{\text{hour}} \cdot 1.25 \text{ hours} = 3.75 \text{ miles}.$$

In addition, we observe³ that if $s(t) = 3t$, then $s'(t) = 3$, so $s(t) = 3t$ is the position function whose derivative is the given velocity function, $v(t) = 3$. The respective locations of the person at times $t = 0.25$ and $t = 1.5$ are $s(1.5) = 4.5$ and $s(0.25) = 0.75$, and therefore

$$s(1.5) - s(0.25) = 4.5 - 0.75 = 3.75 \text{ miles}.$$

This is the person's change in position on $[0.25, 1.5]$, which is precisely the distance traveled. In this example there are profound ideas and connections that we will study throughout Chapter 4.

For now, observe that if we know a formula for a velocity function v , it can be very helpful to find a function s that satisfies $s' = v$. We say that s is an *antiderivative* of v . More generally, we have the following formal definition.

Definition 4.1.9. If g and G are functions such that $G' = g$, we say that G is an **antiderivative** of g .

³Here we are making the implicit assumption that $s(0) = 0$; we will discuss different possibilities for values of $s(0)$ in subsequent study.

For example, if $g(x) = 3x^2 + 2x$, $G(x) = x^3 + x^2$ is an antiderivative of g , because $G'(x) = g(x)$. 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 .

Example 4.1.10. 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 $0 \leq t \leq 2$.

- 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?
- Find an antiderivative, s , of v that satisfies $s(0) = 0$.
- Compute the value of $s(1) - s(\frac{1}{2})$. What is the meaning of the value you find?
- Using the graph of $y = v(t)$ provided in Figure 4.1.11, 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?

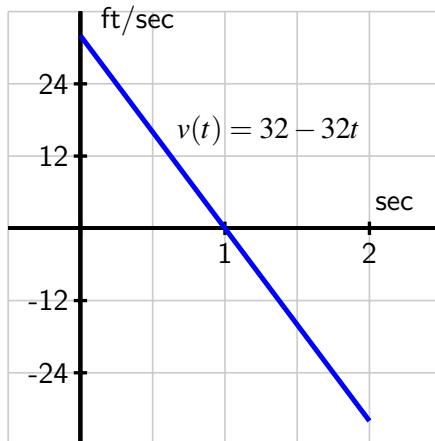


Figure 4.1.11: The graph of $y = v(t)$.

- Answer the same questions as in (c) and (d) but instead using the interval $[0, 1]$.
- 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.

Hint.

- Where is velocity zero?
- Since v is linear, note that s must be quadratic.

- c. Observe that you are taking the difference between two values of the position function.
- d. The region whose area is sought is triangular.
- e. See (c) and (d) above.
- f. What does it mean for the change of the ball's position to be zero?

Answer.

- a. On $(0, 1)$, s is increasing because velocity is positive.
- b. $s(t) = 32t - 16t^2$.
- c. $s(1) - s(\frac{1}{2}) = 4$.
- d. $A = 4$ feet is the total distance the ball traveled vertically on $[0, \frac{1}{2}]$.
- e. $s(1) - s(0) = 16$ is the vertical distance the ball traveled on the interval $[0, 1]$. Equivalently, the area under the velocity curve on $[0, 1]$ is $A = 16$ feet.
- f. $s(2) - s(0) = 0$, so the ball has zero change in position on the interval $[0, 2]$.

Solution.

- a. Note that $v(1) = 0$ and for $0 < t < 1$, $v(t) > 0$. This means that on the interval $(0, 1)$, the position function s is increasing because velocity is positive.
- b. We can check that the derivative of $s(t) = 32t - 16t^2$ is $s'(t) = v(t) = 32 - 32t$, and that $s(0) = 0$, so this is the antiderivative of v that we desire.
- c. Now, $s(1) - s(\frac{1}{2}) = (32 - 16) - (16 - 4) = 4$, which is the change in position of the ball on the interval $[\frac{1}{2}, 1]$. Equivalently, since v is positive through this interval, 4 feet is the vertical distance the ball traveled during this time.
- d. On the interval from $t = \frac{1}{2}$ to $t = 1$, the corresponding area under the velocity curve is the area of the right triangular region whose width is $\frac{1}{2}$ seconds and whose height is $v(\frac{1}{2}) = 16$ feet/sec. That area is therefore $A = \frac{1}{2}bh = \frac{1}{2} \cdot \frac{1}{2} \cdot 16 = 4$ feet. This is the total distance the ball traveled vertically on $[0, \frac{1}{2}]$.
- e. $s(1) - s(0) = (32 - 16) - (0 - 0) = 16$, which is the vertical distance the ball traveled on the interval $[0, 1]$. The area under the velocity curve on $[0, 1]$ is the area of the triangle with height 32 (ft/sec) and base 1 (second), which is $A = \frac{1}{2} \cdot 1 \cdot 32 = 16$ feet. These two results are identical, in part due to the fact that we are using two different perspectives to compute the same quantity, which is distance traveled.
- f. Observe that $s(2) - s(0) = (32 - 32) - (0 - 0) = 0$. This means that the ball has zero change in position on the interval $[0, 2]$. But we already established that on the interval $[0, 1]$, the ball traveled 16 feet vertically; since the velocity becomes negative on the interval $1 < t < 2$, there we know the ball's position is decreasing, so it is falling back to earth. The resulting zero change in position means that at $t = 2$ the ball has returned to the

location from which it was tossed. If we view the area between the velocity function and the t -axis as being negative wherever v is negative, then we see that the areas of the two triangles involved are opposites, which in some sense results in the “total area” being zero, matching the change in position.

4.1.3 When velocity is negative

The assumption that its velocity is positive on a given interval guarantees that the movement of an object is always in a single direction, and hence ensures that its change in position is the same as the distance it travels. As we saw in Example 4.1.10, there are natural settings in which an object’s velocity is negative, and we would like to understand this scenario as well.

Consider a simple example where a woman goes for a walk on the beach along a stretch of very straight shoreline that runs east-west. We assume that her initial position is $s(0) = 0$, and that her position function increases as she moves east from her starting location. For instance, $s = 1$ mile represents one mile east of the start location, while $s = -1$ tells us she is one mile west of where she began walking on the beach.

Now suppose she walks in the following manner. From the outset at $t = 0$, she walks due east at a constant rate of 3 mph for 1.5 hours. After 1.5 hours, she stops abruptly and begins walking due west at a constant rate of 4 mph and does so for 0.5 hours. Then, after another abrupt stop and start, she resumes walking at a constant rate of 3 mph to the east for one more hour. What is the total distance she traveled on the time interval from $t = 0$ to $t = 3$? What is the total change in her position over that time?

These questions are possible to answer without calculus because the velocity is constant on each interval. From $t = 0$ to $t = 1.5$, she traveled

$$D_{[0,1.5]} = 3 \text{ miles per hour} \cdot 1.5 \text{ hours} = 4.5 \text{ miles.}$$

On $t = 1.5$ to $t = 2$, the distance traveled is

$$D_{[1.5,2]} = 4 \text{ miles per hour} \cdot 0.5 \text{ hours} = 2 \text{ miles.}$$

Finally, in the last hour she walked

$$D_{[2,3]} = 3 \text{ miles per hour} \cdot 1 \text{ hours} = 3 \text{ miles,}$$

so the total distance she 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 for $1.5 < t < 2$ is $v = -4$, indicating motion in the westward direction, the woman first walked 4.5 miles east, then 2 miles west, followed by 3 more miles east. Thus, the total change in her position is

$$\text{change in position} = 4.5 - 2 + 3 = 5.5 \text{ miles.}$$

We have been able to answer these questions fairly easily, and if we think about the problem graphically, we can generalize our solution to the more complicated setting when velocity is not constant, and possibly negative.

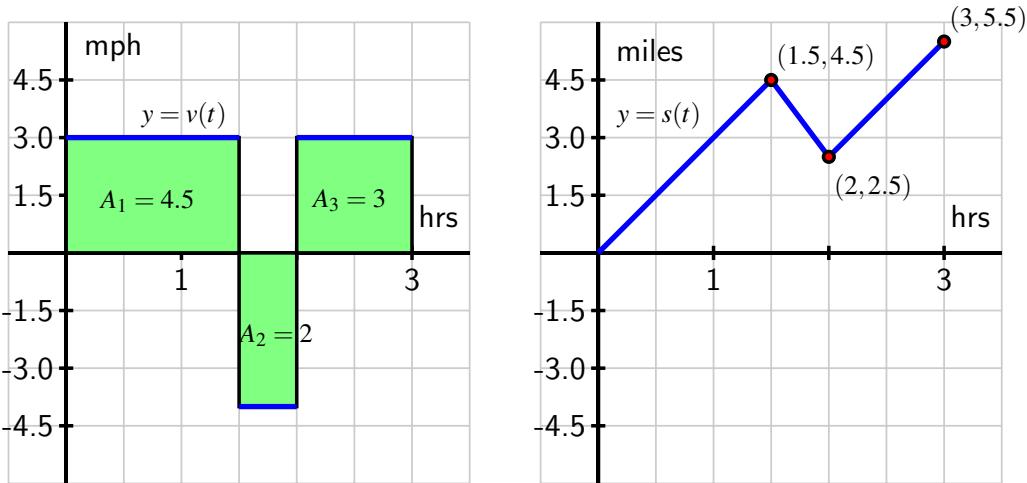


Figure 4.1.12: At left, the velocity function of the person walking; at right, the corresponding position function.

In Figure 4.1.12, we see how the distances we computed can be viewed as areas: $A_1 = 4.5$ comes from multiplying rate times time ($3 \cdot 1.5$), as do A_2 and A_3 . But while A_2 is an area (and is therefore positive), because the velocity function is negative for $1.5 < t < 2$, this area has a negative sign associated with it. The negative area distinguishes 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 travel in the negative direction. An area above the t -axis is considered positive because it represents distance traveled in the positive direction, while one below the t -axis is viewed as negative because it represents travel in the negative direction. Thus, the change in the woman's position is

$$s(3) - s(0) = (+4.5) + (-2) + (+3) = 5.5 \text{ miles.}$$

In other words, the woman 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.

Negative velocity is also seen in the graph of the position function $y = s(t)$. Its slope is negative (specifically, -4) on the interval $1.5 < t < 2$ because the velocity is -4 on that interval. The negative slope shows the position function is decreasing because the woman is walking east, rather than west.

To summarize, we see that if velocity is sometimes negative, a moving object's change in po-

sition different from its distance traveled. If we compute separately the distance traveled on each interval where velocity is positive or negative, we can calculate either the total distance traveled or the total change in position. We assign a negative value to distances traveled in the negative direction when we calculate change in position, but a positive value when we calculate the total distance traveled.

Example 4.1.13. 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 at left in Figure 4.1.14. 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.

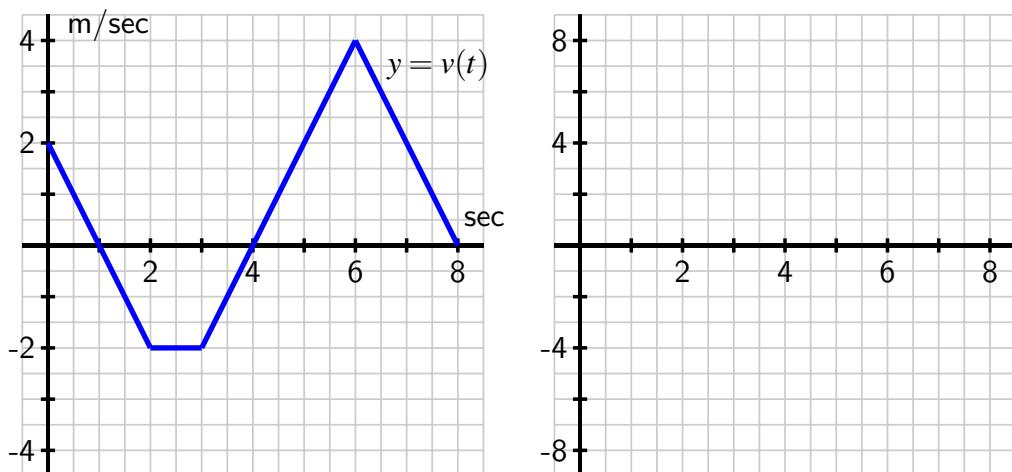


Figure 4.1.14: The velocity function of a moving object.

Suppose further that the object's initial position at time $t = 0$ is $s(0) = 1$.

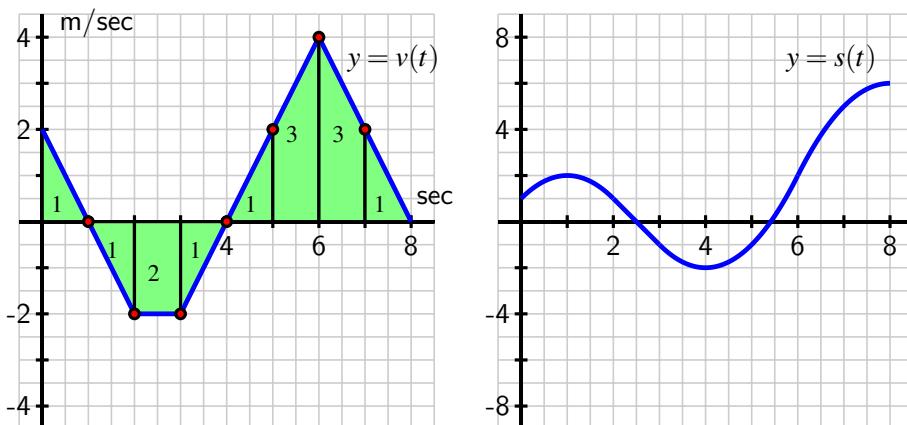
- Determine the total distance traveled and the total change in position on the time interval $0 \leq t \leq 2$. What is the object's position at $t = 2$?
- On what time intervals is the moving object's position function increasing? Why? On what intervals is the object's position decreasing? Why?
- What 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 $t = 8$?
- Find the exact position of the object at $t = 1, 2, 3, \dots, 8$ and use this data to sketch an accurate graph of $y = s(t)$ on the axes provided at right in Figure 4.1.14. How can you use the provided information about $y = v(t)$ to determine the concavity of s on each relevant interval?

Hint.

- Find the area of each triangular region formed between $y = v(t)$ and the t -axis.
- Recall that $v = s'$, and here we are given complete information about v .
- Be careful to address whether v is positive or negative when calculating areas and adding the results.
- Consider finding the area bounded by $y = v(t)$ and the t -axis on each interval $[0, 1]$, $[1, 2]$, ...

Answer.

- Total distance traveled is 2; change in position is 0.
- $0 < t < 1$ and $4 < t < 8$.
- $s(8) - s(0) = 1 - 4 + 8 = 5$ m.
- See the figure below.



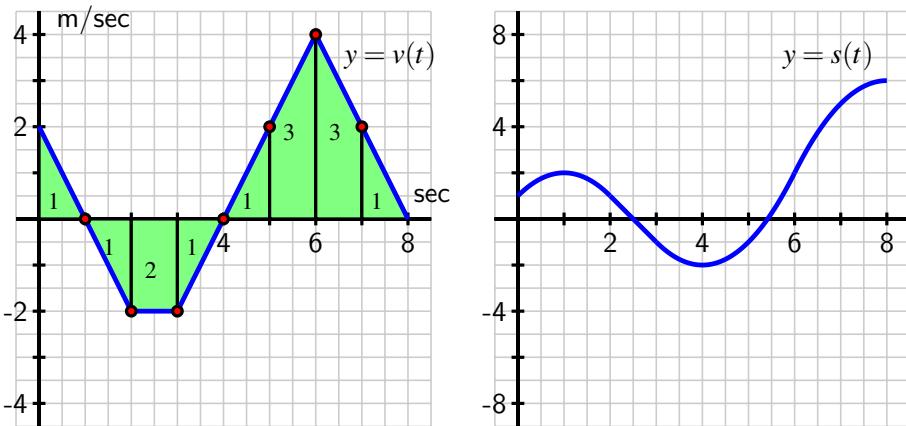
Solution.

- By finding the area of the triangular regions formed between $y = v(t)$ and the t -axis on $[0, 1]$ and $[1, 2]$ (each of which is 1), it follows that the object's total distance traveled is 2, while its change in position is 0. The latter is true since the net signed area bounded by v on $[0, 2]$ is $1 - 1 = 0$.
- The object's position is increasing wherever its velocity is positive, hence for $0 \leq t < 1$ and $4 < t < 8$.
- By calculating the area bounded by the curve, we find 1 unit of area on $[0, 1]$, 4 units of area on $[1, 4]$, and 8 units of area on $[4, 8]$, thus the total distance traveled on $0 \leq t \leq 8$

is $D = 1 + 4 + 8$ meters. As the change in position is given by the net signed area on this interval, we find that the change in position is

$$s(8) - s(0) = 1 - 4 + 8 = 5 \text{ m.}$$

- d. In the figure below, at left we list all of the areas bounded by v on each one-unit subinterval. Along with the given starting point that $s(0) = 1$, we use the resulting changes in position to plot points for the function s . For instance, we know $s(1) - s(0) = 1$, hence $s(1) = 2$. Similarly, $s(2) - s(1) = -1$, thus $s(2) = 1$. Continuing across the interval, we generate the function s that is pictured at right. Note that the portion of s from $t = 2$ to $t = 3$ is linear because v is constant there, while the other parts of s appear to be quadratic, as they correspond to intervals where v is linear.



4.1.4 Summary

- 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.
- An 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 $[a, b]$.

- If its velocity is sometimes negative, a moving object is sometimes traveling in the opposite direction or backtracking. To determine distance traveled, we have to think compute the distance separately on intervals where velocity is positive or negative, and account for the change in position on each such interval.

4.1.5 Exercises

1. (Estimating distance traveled from velocity data) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Distance from a linear velocity function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Change in position from a linear velocity function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Comparing distance traveled from velocity graphs) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Finding average acceleration from velocity data) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Change in position from a quadratic velocity function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. 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 in Figure 4.1.15 on the time interval $0 \leq t \leq 4$ (where t is measured in hours), and that $s(0) = 1$.

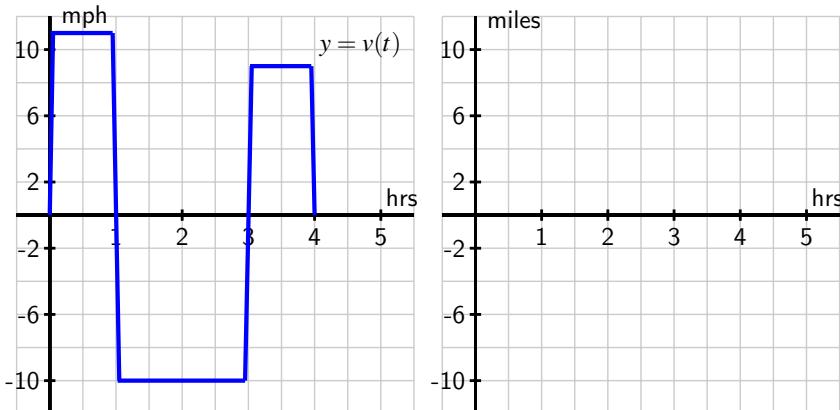
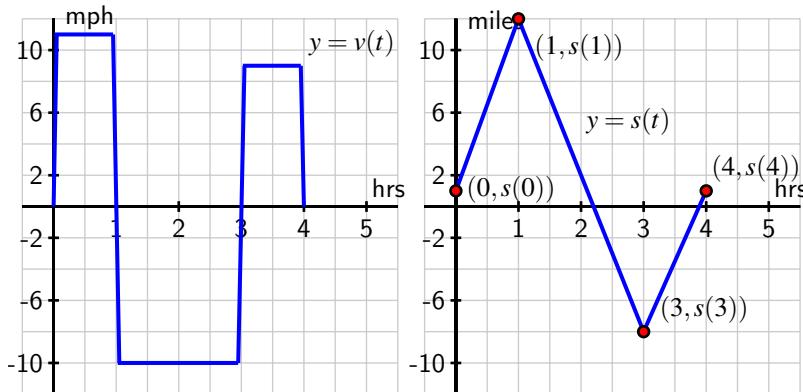


Figure 4.1.15: The graph of the biker's velocity, $y = v(t)$, at left. At right, axes to plot an approximate sketch of $y = s(t)$.

- 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?
- What is the cyclist's total change in position on the time interval $0 \leq t \leq 2$? At $t = 2$, was she north or south of Pigeon Lake?
- What is the total distance the biker traveled on $0 \leq t \leq 4$? At the end of the ride, how close was she to the point at which she started?
- Sketch an approximate graph of $y = s(t)$, the position function of the cyclist, on the interval $0 \leq t \leq 4$. Label at least four important points on the graph of s .

Answer.

- At time $t = 1$, 12 miles north of the lake.
- $s(4) - s(0) = 0$.
- 40 miles.
-



Solution.

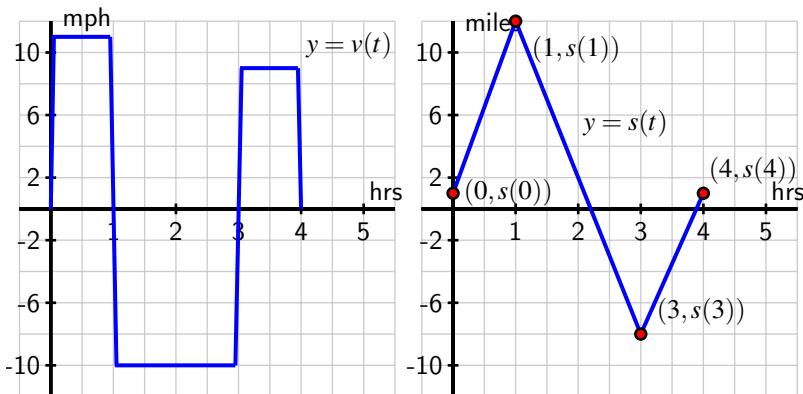
- The velocity function of the biker tells us that for the first hour, the biker pedaled north at a velocity of 11 miles per hour. At the 1-hour mark, the biker abruptly changes direction and then pedaled south at 10 miles per hour for the next two hours; finally, at the 3-hour mark, the biker again changes direction and then pedaled north at 9 mph for the last hour of their trip.

Thus, the biker traveled 11 miles north from their starting point (mile 1) and thus made the turn at $t = 1$ at mile 12. Two hours later, having pedaled south at 10 mph, they'd traveled 20 miles south, and thus were at $12 - 20 = -8$ relative to Pigeon Lake. Finally, after biking at 9 mph north for the last hour, they were back at $-8 + 9 = 1$.

The furthest north the biker was occurred at time $t = 1$. This occurs because it is the only time when the derivative of the biker's position function changes from positive to negative, and thus s has a local maximum. In addition, we can see this based on our discussion above of how far the biker traveled in each direction on the various time intervals, and that the biker was 12 miles north of the lake at $t = 1$.

- The cyclist's total change in position is $s(4) - s(0) = 0$. In part (a), we established that $s(1) = 12$, $s(3) = -8$, and $s(4) = 1$. We were given that $s(0) = 1$. Thus, $s(4) - s(0) = 0$, and the biker ended up back where they started the trip.

- c. The cyclist first traveled 11 miles north, then 20 miles south, and finally 9 miles north. Since $11 + 20 + 9 = 40$, their total distance traveled is 40 miles, even though their change of position 0, as established in (b).
- d. In the following figure, we see an approximate plot of the biker's position function. Since their velocity is constant on the various intervals, their position function is linear on those intervals with slope that matches the velocity. In addition, as pictured, the position function passes through the points we've established above like $(3, -8)$.



8. 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.

- 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 $t = a$?
- Find the value of the total area enclosed by $y = v(t)$ and the t -axis on the interval $0 \leq t \leq a$. What does this area represent in terms of the physical setting of the problem?
- Find an antiderivative s of the function v . That is, find a function s such that $s'(t) = v(t)$.
- Compute the value of $s(a) - s(0)$. What does this number represent in terms of the physical setting of the problem?
- Compute $s(5) - s(1)$. What does this number tell you about the rocket's flight?

Answer.

- $t = \frac{500}{32} = \frac{125}{8} = 15.625$ is when the rocket reaches its maximum height.
- $A = 3906.25$, the vertical distance traveled on $[0, 15.625]$.
- $s(t) = 500t - 16t^2$.
- $s(15.625) - s(0) = 3906.25$ is the change of the rocket's position on $[0, 15.625]$.
- $s(5) - s(1) = 1616$; the rocket rose 1616 feet on $[1, 5]$.

Solution.

- Setting $v(t) = 0$, we find $500 - 32t = 0$, and thus $t = \frac{500}{32} = \frac{125}{8} = 15.625$. This is the time that the rocket's velocity changes from positive to negative, and thus the point when the rocket's position changes from increasing to decreasing: it's when the rocket reaches its maximum height.

- b. Since v is a positive linear function that connects the points $(0, 500)$ and $(15.625, 0)$ on the interval $[0, 15.625]$, the region enclosed by the velocity function and the t -axis is a right triangle with vertical leg of length 500 and horizontal leg of length 15.625. The area of this triangle is $A = \frac{1}{2}bh = \frac{1}{2}(15.625)(500) = 3906.25$. Because this is the area under the velocity curve (where velocity is positive) this area represents the vertical distance traveled by the rocket on the time interval $[0, 15.625]$.
- c. One function that has derivative $v(t) = s'(t) = 500 - 32t$ is $s(t) = 500t - 16t^2$.
- d. Using our result in (c), $s(15.625) - s(0) = 500(15.625) - 16(15.625)^2 - (0 - 0) = 3906.25$. Thus, the change of the rocket's position on $[0, 15.625]$ is 3906.25 feet, which is also the distance the rocket traveled vertically on the time interval, since velocity is positive throughout.
- e. We find that $s(5) - s(1) = 1616$, which tells us that the rocket rose 1616 feet on the 4-second interval from $t = 1$ to $t = 5$.
9. An object moving along a horizontal axis has its instantaneous velocity at time t in seconds given by the function v pictured in Figure 4.1.16, 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.

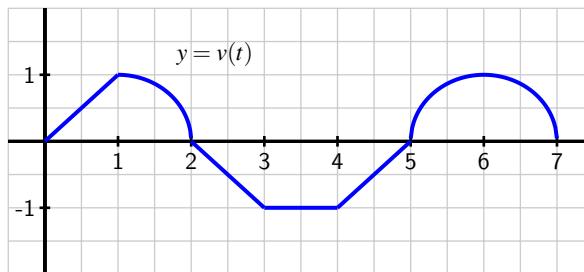


Figure 4.1.16: The graph of $y = v(t)$, the velocity function of a moving object.

- a. Determine the exact total distance the object traveled on $0 \leq t \leq 2$.
- 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?

Answer.

- a. $\frac{1}{2} + \frac{1}{4}\pi \approx 1.285$.
- b. $s(5) - s(2) = -2$ is the change in position of the object on $[2, 5]$.
- c. On the time interval $[5, 7]$.
- d. s is increasing on the intervals $(0, 2)$ and $(5, 7)$; the position function has a relative maximum at $t = 2$.

Solution.

- a. To find the exact total distance traveled, we need to calculate the area under the velocity function on the interval $[0, 2]$. This area is the sum of the areas of the triangular region of height 1 and base length 1 on the interval $[0, 1]$ plus the area of the quarter circle of radius 1 on the interval $[1, 2]$. So the total distance the object traveled on $0 \leq t \leq 2$ is

$$\frac{1}{2}(1)(1) + \frac{1}{4}\pi(1^2) = \frac{1}{2} + \frac{1}{4}\pi \approx 1.285.$$

- b. The meaning of $s(5) - s(2)$ is the change in position, or total displacement, of the object on the time interval $[2, 5]$. Note that $s(2)$ was found to be $\frac{1}{2} + \frac{1}{4}\pi$ in part (a) and

$$s(5) = \frac{1}{2} + \frac{1}{4}\pi - \frac{1}{2} - 1 - \frac{1}{2} = \frac{\pi}{4} - \frac{3}{2}.$$

So, $s(5) - s(2) = -2$, which means that the object moved 2 feet back toward its original position from time $t = 2$ to time $t = 5$.

- c. In part (a) we calculated the distance traveled by the object on the interval $[0, 2]$ to be approximately 1.285 feet. On the time interval $[2, 4]$, the object is moving in the opposite direction from the motion on the interval $[0, 2]$, and the total distance traveled is the area $v(t)$ bounds below the t -axis on the interval $[2, 4]$. This area is the sum of the areas of the triangle of height 1 and base length 1 on the interval $[2, 3]$ and the area of the rectangle of height 1 and length 1 on the interval $[3, 4]$. So the total distance traveled by the object on the interval $[2, 4]$ is

$$\frac{1}{2}(1)(1) + (1)(1) = 1.5.$$

The total distance traveled by the object on the interval $[5, 7]$ is the area $v(t)$ bounds above the t -axis on the interval $[5, 7]$. This area is the area of the semicircle of radius 1 on the interval $[5, 7]$. So the total distance traveled by the object on the interval $[5, 7]$ is

$$\frac{1}{2}\pi(1)^2 = \frac{\pi}{2} \approx 1.571.$$

Thus, the object traveled the greatest distance on the time interval $[5, 7]$.

- d. We know that the position function is increasing when the velocity is positive, so s is increasing on the intervals $(0, 2)$ and $(5, 7)$. Since $v(t)$ is positive to the left of $t = 2$ and then negative to the right of $t = 2$, the position function s is increasing to the left of $t = 2$ then decreasing to the right of $t = 2$. Thus, the position function has a relative maximum point at $t = 2$. There are no other points with this property, so this is the only relative maximum value of s on this interval. We also can note that s increases from $t = 5$ to $t = 7$, culminating in a local maximum value at $t = 7$.

10. 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 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.

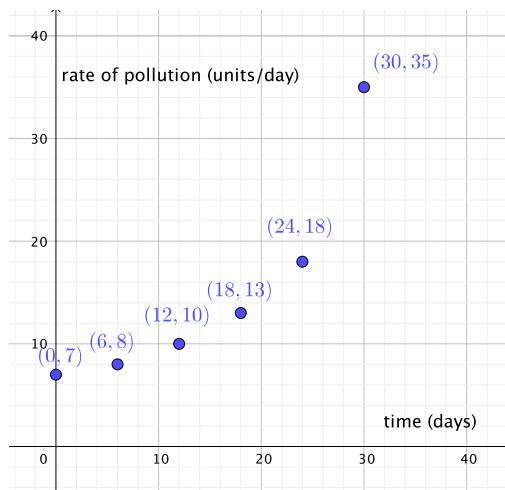
Day, t	0	6	12	18	24	30
Rate of pollution in units per day, $p(t)$	7	8	10	13	18	35

Table 4.1.17: Pollution data for the water filters.

- Plot the given data on a set of axes with time on the horizontal axis and the rate of pollution on the vertical axis.
- 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 $[0, 30]$.
- Estimate the total amount of pollution entering the lake during this 30-day period. Carefully explain how you determined your estimate.

Answer.

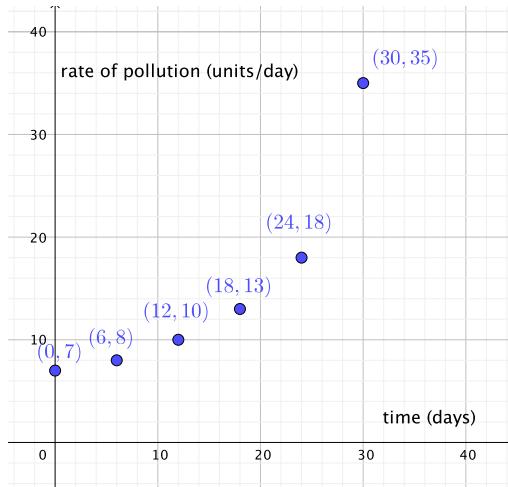
a.



- Think about the product of the units involved: “units of pollution per day” times “days”. Connect this to the area of a thin vertical rectangle whose height is given by the curve.
- An underestimate is 546 units of pollution.

Solution.

- The plot below shows that the rate of pollution appears to be increasing at an increasing rate as time goes on.



- b. If we think about an interval on which the rate of pollution is relatively constant and suppose the interval is Δt wide, then the area under the curve on that interval is approximately the area of the rectangle whose width is Δt and whose height is $p(t)$, the pollution rate on that interval. When we compute the product $p(t) \cdot \Delta t$ to find the area, the units involved are “units of pollution per day” times “days”, which results in an overall unit of “units of pollution”. Thus, the area under the curve tells us the total pollution emitted over the time interval $[0, 30]$.
- c. One way to estimate the total amount of pollution is to suppose the rate of pollution is constant on each of the intervals $[0, 6]$, $[6, 12]$, For instance, we could assume the rate on $[0, 6]$ is always 7 units per day, and then the rate jumps to 8 units per day on $[6, 12]$, and so on. Doing so, and using the ideas we discussed in (b), one estimate for the total pollution, P , is

$$P = 7 \cdot 6 + 8 \cdot 6 + 10 \cdot 6 + 13 \cdot 6 + 18 \cdot 6 + 35 \cdot 6 = 546$$

total units of pollution. Note, for example, that when we compute $13 \cdot 6$ this represents an estimate of the total amount of pollution emitted on the interval $[18, 24]$ since we are assuming the plant emits 13 units of pollution per day for 6 days. It is apparent that this particular estimate is an underestimate of the total amount of pollution since the rate is always increasing but we are choosing the lowest value of that rate on a given interval.

4.2 Riemann Sums

Motivating 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?
- How can we write Riemann sums in an abbreviated form?

In Section 4.1, we learned that if an object moves with positive velocity v , 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. 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 tells us the moving object's change in position over a given time interval.

For instance, for the velocity function given in Figure 4.2.1, 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

$$s(b) - s(a) = A_1 - A_2 + A_3.$$

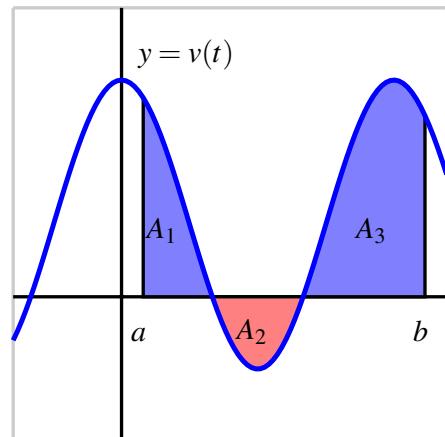


Figure 4.2.1: A velocity function that is sometimes negative.

Because the motion is in the negative direction on the interval where $v(t) < 0$, we subtract A_2 to determine the object's total change in position.

Of course, finding D and $s(b) - s(a)$ for the graph in Figure 4.2.1 presumes that we can actually find the areas A_1 , A_2 , and A_3 . So far, we have worked with velocity functions that were either constant or linear, so that the area bounded by the velocity function and the horizontal axis is a combination of rectangles and triangles, and we can find the area exactly. But when the curve bounds a region that is not a familiar geometric shape, we cannot find its 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 the horizontal axis for as many different types of functions as possible.

In Example 4.1.5, we approximated the area under a nonlinear velocity function using rectangles. In Example 4.2.2, we consider three different options for the heights of the rectangles we will use.

Example 4.2.2. 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 in Figure 4.2.3.

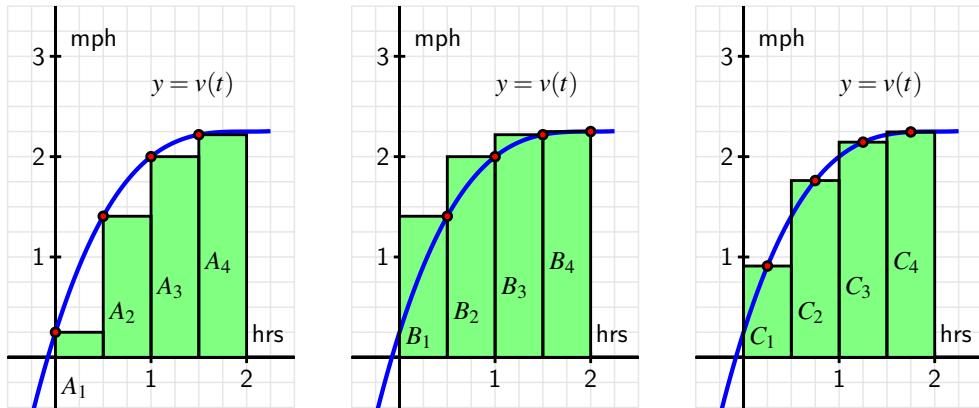


Figure 4.2.3: Three approaches to estimating the area under $y = v(t)$ on the interval $[0, 2]$.

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?

4.2.1 Sigma Notation

We have used sums of areas of rectangles to approximate the area under a curve. Intuitively, we expect that using a larger number of thinner rectangles will provide a better estimate for the area. Consequently, we anticipate dealing with sums of a large number of terms. To do so, we introduce *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

$$1 + 2 + 3 + \cdots + 100,$$

the sum of the first 100 natural numbers. In sigma notation we write

$$\sum_{k=1}^{100} k = 1 + 2 + 3 + \cdots + 100.$$

We read the symbol $\sum_{k=1}^{100} k$ as “the sum from k equals 1 to 100 of k .” The variable k is called the index of summation, and any letter can be used for this variable. The pattern in the terms of the sum is denoted by a function of the index; for example,

$$\sum_{k=1}^{10} (k^2 + 2k) = (1^2 + 2 \cdot 1) + (2^2 + 2 \cdot 2) + (3^2 + 2 \cdot 3) + \cdots + (10^2 + 2 \cdot 10),$$

and more generally,

$$\sum_{k=1}^n f(k) = f(1) + f(2) + \cdots + f(n).$$

Sigma notation allows us to vary easily the function being used to describe the terms in the sum, and 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 example.

Example 4.2.4. 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. $\sum_{k=1}^5 (k^2 + 2)$ | d. $4 + 8 + 16 + 32 + \cdots + 256$ |
| b. $\sum_{i=3}^6 (2i - 1)$ | |
| c. $3 + 7 + 11 + 15 + \cdots + 27$ | e. $\sum_{i=1}^6 \frac{1}{2^i}$ |

Hint.

- a. Observe that when $k = 1$, $k^2 + 2 = 1^2 + 2 = 3$. This is the first term in the sum.

- b. Note that this sum starts at $i = 3$.
- c. Since the terms in the sum increase by 4, try a function $f(k)$ that somehow involves $4k$.
- d. What pattern do you observe in the terms of the sum?
- e. Write every term in the sum as a fraction with denominator $2^6 = 64$.

Answer.

a. 64

b. 32

c.

$$3 + 7 + 11 + 15 + \cdots + 27 = \sum_{k=1}^7 4k - 1.$$

d.

$$4 + 8 + 16 + 32 + \cdots + 256 = \sum i = 2^8 2^i.$$

e.

$$\sum_{i=1}^6 \frac{1}{2^i} = \frac{63}{64}.$$

Solution.

a.

$$\begin{aligned} \sum_{k=1}^5 (k^2 + 2) &= (1^2 + 2) + (2^2 + 2) + (3^2 + 2) + (4^2 + 2) + (5^2 + 2) \\ &= 3 + 6 + 11 + 17 + 27 \\ &= 64 \end{aligned}$$

b.

$$\begin{aligned} \sum_{i=3}^6 (2i - 1) &= (2 \cdot 3 - 1) + (2 \cdot 4 - 1) + (2 \cdot 5 - 1) + (2 \cdot 6 - 1) \\ &= 5 + 7 + 9 + 11 \\ &= 32 \end{aligned}$$

c. Observe that each term in the sum

$$3 + 7 + 11 + 15 + \cdots + 27$$

differs from the previous term by 4. If we view 4 as $4 = 4 \cdot 1 - 1$ and 7 as $7 = 4 \cdot 2 - 1$, we see that the pattern may be represented through the function $f(k) = 4k - 1$, so that

$$3 + 7 + 11 + 15 + \cdots + 27 = \sum_{k=1}^7 4k - 1.$$

We note that $k = 7$ is the end value of the index since $4 \cdot 7 = 28$.

- d. The sum $4 + 8 + 16 + 32 + \cdots + 256$ is a sum of powers of 2, which we can express in sigma notation as

$$4 + 8 + 16 + 32 + \cdots + 256 = \sum i = 2^8 2^i.$$

e.

$$\begin{aligned} \sum_{i=1}^6 \frac{1}{2^i} &= \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^6} \\ &= \frac{32}{64} + \frac{16}{64} + \frac{8}{64} + \frac{4}{64} + \frac{2}{64} + \frac{1}{64} \\ &= \frac{63}{64}. \end{aligned}$$

4.2.2 Riemann Sums

When a moving body has a positive velocity function $y = v(t)$ on a given interval $[a, b]$, the area under the curve over the interval gives the total distance the body travels on $[a, b]$. We are also interested in finding 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 finding an accurate estimate of this area by using a sum of the areas of rectangles. Unless otherwise indicated, we assume that f is continuous and non-negative on $[a, b]$.

The first choice we make in such an approximation is the number of rectangles.

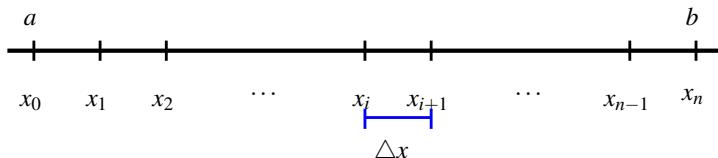


Figure 4.2.5: Subdividing the interval $[a, b]$ into n subintervals of equal length Δx .

If we desire n rectangles of equal width to subdivide the interval $[a, b]$, then each rectangle must have width $\Delta x = \frac{b-a}{n}$. We let $x_0 = a$, $x_n = b$, and define $x_i = a + i\Delta x$, so that $x_1 = x_0 + \Delta x$, $x_2 = x_0 + 2\Delta x$, and so on, as pictured in Figure 4.2.5.

We use each subinterval $[x_i, x_{i+1}]$ as the base of a rectangle, and next choose the height of the rectangle on that subinterval. There are three standard choices: we can 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 Example 4.2.2 and seen in Figure 4.2.3. We next explore how these choices can be described in sigma notation.

Consider an arbitrary positive function f on $[a, b]$ with the interval subdivided as shown in Figure 4.2.5, and choose to use left endpoints. Then on each interval $[x_i, x_{i+1}]$, the area of the rectangle formed is given by

$$A_{i+1} = f(x_i) \cdot \Delta x,$$

as seen in Figure 4.2.6.

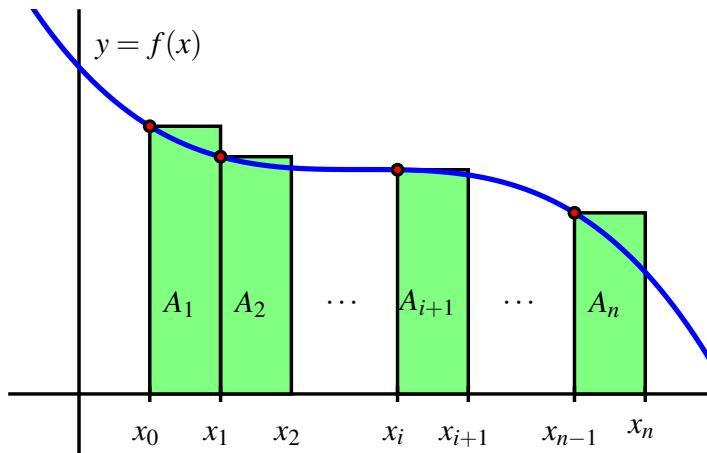


Figure 4.2.6: Subdividing the interval $[a, b]$ into n subintervals of equal length Δx and approximating the area under $y = f(x)$ over $[a, b]$ using left rectangles.

If we let L_n denote the sum of the areas of these rectangles, we see that

$$\begin{aligned} L_n &= A_1 + A_2 + \cdots + A_{i+1} + \cdots + A_n \\ &= f(x_0) \cdot \Delta x + f(x_1) \cdot \Delta x + \cdots + f(x_i) \cdot \Delta x + \cdots + f(x_{n-1}) \cdot \Delta x. \end{aligned}$$

In the more compact sigma notation, we have

$$L_n = \sum_{i=0}^{n-1} f(x_i) \Delta x.$$

Note 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 $[a, b]$.

To see how the Riemann sums for right endpoints and midpoints are constructed, we consider Figure 4.2.7.

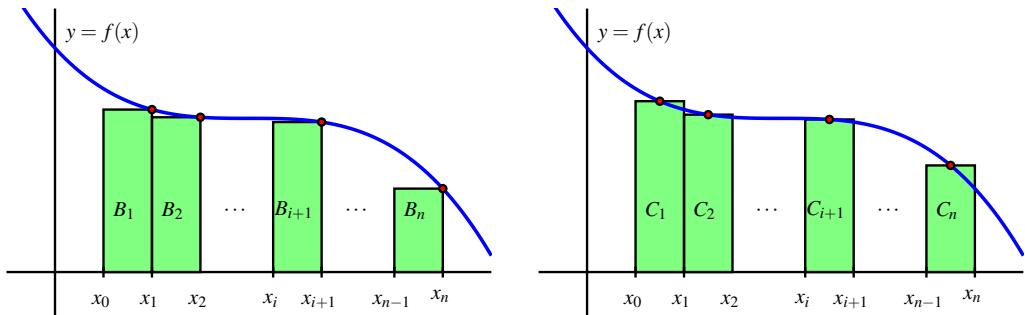


Figure 4.2.7: Riemann sums using right endpoints and midpoints.

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 $B_{i+1} = f(x_{i+1}) \cdot \Delta x$, and that the sum of all such areas of rectangles is given by

$$\begin{aligned} R_n &= B_1 + B_2 + \cdots + B_{i+1} + \cdots + B_n \\ &= f(x_1) \cdot \Delta x + f(x_2) \cdot \Delta x + \cdots + f(x_{i+1}) \cdot \Delta x + \cdots + f(x_n) \cdot \Delta x \\ &= \sum_{i=1}^n f(x_i) \Delta x. \end{aligned}$$

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

$$\bar{x}_{i+1} = \frac{x_i + x_{i+1}}{2}$$

so that \bar{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 4.2.7, we now have

$$C_1 = f(\bar{x}_1) \cdot \Delta x.$$

Hence, the sum of all the areas of rectangles that use midpoints is

$$\begin{aligned} M_n &= C_1 + C_2 + \cdots + C_{i+1} + \cdots + C_n \\ &= f(\bar{x}_1) \cdot \Delta x + f(\bar{x}_2) \cdot \Delta x + \cdots + f(\bar{x}_{i+1}) \cdot \Delta x + \cdots + f(\bar{x}_n) \cdot \Delta x \\ &= \sum_{i=1}^n f(\bar{x}_i) \Delta x, \end{aligned}$$

and we say that M_n is the *middle Riemann sum* for f on $[a, b]$.

Thus, we have two variables to explore: the number of rectangles and the height of each rectangle. We can explore these choices dynamically, and the applet¹ found at <http://gvsu.edu/s/a9> is a particularly useful one. There we see the image shown in Figure 4.2.8, but with the opportunity to adjust the slider bars for the heights and the number of rectangles.

¹Marc Renault, Geogebra Calculus Applets.

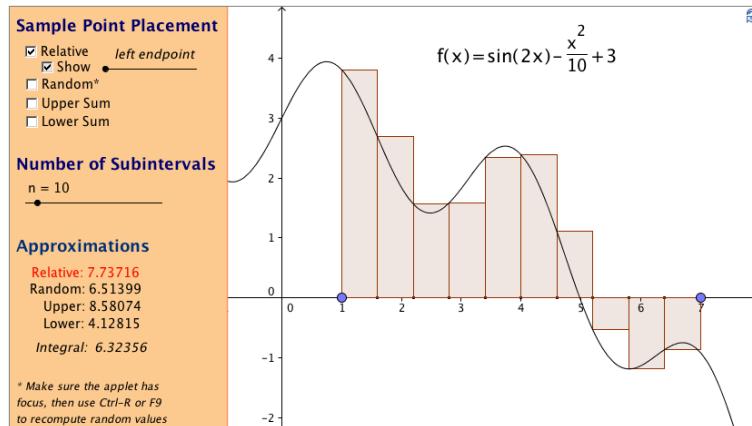


Figure 4.2.8: A snapshot of the applet found at <http://gvsu.edu/s/a9>.

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.

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 $[a, b]$. We also recall that in the context of a nonnegative velocity function $y = v(t)$, the corresponding Riemann sums approximate the distance traveled on $[a, b]$ by a moving object with velocity function v .

There is a more general way to think of Riemann sums, and that is to allow any choice of where the function is evaluated to determine the rectangle heights. 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 $x_i \leq x_{i+1}^* \leq x_{i+1}$). The Riemann sum is then given by

$$f(x_1^*) \cdot \Delta x + f(x_2^*) \cdot \Delta x + \cdots + f(x_{i+1}^*) \cdot \Delta x + \cdots + f(x_n^*) \cdot \Delta x = \sum_{i=1}^n f(x_i^*) \Delta x.$$

At <http://gvsu.edu/s/a9>, the applet noted earlier and referenced in Figure 4.2.8, 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 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 example, we investigate several different Riemann sums for a particular velocity function.

Example 4.2.9. Suppose that an object moving along a straight line path has its velocity in feet per second at time t in seconds given by $v(t) = \frac{2}{9}(t - 3)^2 + 2$.

- Carefully sketch the region whose exact area will tell you the value of the distance the object traveled on the time interval $2 \leq t \leq 5$.
- Estimate the distance traveled on $[2, 5]$ by computing L_4 , R_4 , and M_4 .

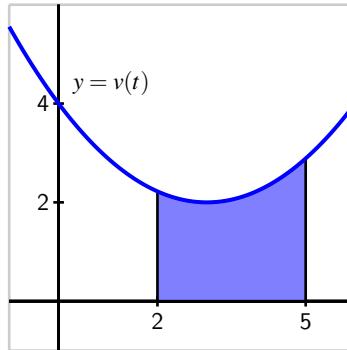
- c. Does averaging L_4 and R_4 result in the same value as M_4 ? If not, what do you think the average of L_4 and R_4 measures?
- d. For this question, think about an arbitrary function f , rather than the particular function v given above. If f is positive and increasing on $[a, b]$, will L_n over-estimate 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.

Hint.

- a. Note that $y = v(t)$ is a parabola with vertex $(3, 2)$.
- b. Recall the formulas for L_n , R_n , and M_n .
- c. Think about what the average of L_1 and R_1 measures.
- d. Consider carefully the role of endpoints in generating L_n and R_n .

Answer.

a.



b. $L_4 = \frac{311}{48} \approx 6.47917$, $R_4 = \frac{335}{48} \approx 6.97917$, and $M_4 = \frac{637}{96} \approx 6.63542$.

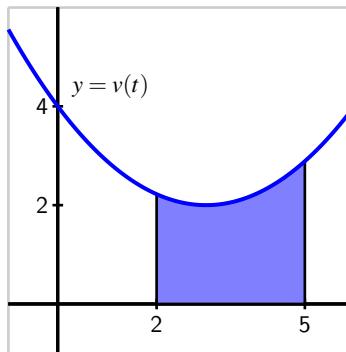
c.

$$\frac{L_4 + M_4}{2} = \frac{646}{96} \neq \frac{637}{96} = M_4.$$

d. L_n is an under-estimate; R_n is an over-estimate.

Solution.

- a. The region whose exact area tells us the value of the distance the object traveled on the time interval $2 \leq t \leq 5$ is shown below.



b. $L_4 = \frac{311}{48} \approx 6.47917$, $R_4 = \frac{335}{48} \approx 6.97917$, and $M_4 = \frac{637}{96} \approx 6.63542$.

c. The average of L_4 and R_4 is

$$\frac{L_4 + M_4}{2} = \frac{311 + 335}{96} = \frac{646}{96} \neq \frac{637}{96} = M_4.$$

This average actually measures what would result from using four trapezoids, rather than rectangles, to estimate the area on each subinterval. One reason this is so is because the area of a trapezoid is the average of the bases times the width, and the “bases” are given by the function values at the left and right endpoints.

d. If f is positive and increasing on $[a, b]$, L_n will under-estimate the exact area under f on $[a, b]$. Because f is increasing, its value at the left endpoint of any subinterval will be lower than every other function value in the interval, and thus the rectangle with that height lies exclusively below the curve. In a similar way, R_n over-estimates the exact area under f on $[a, b]$.

4.2.3 When the function is sometimes negative

For a Riemann sum such as

$$L_n = \sum_{i=0}^{n-1} f(x_i) \Delta x,$$

we can of course compute the sum even when f takes on negative values. We know that when f is positive on $[a, b]$, a Riemann sum estimates the area bounded between f and the horizontal axis over the interval.

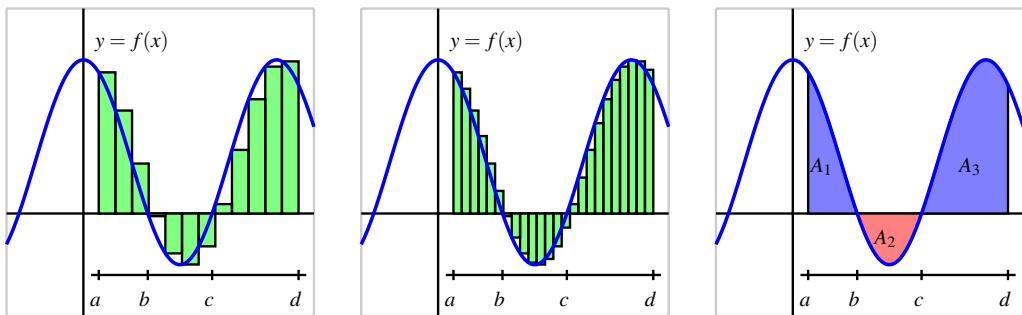


Figure 4.2.10: At left and center, two left Riemann sums for a function f that is sometimes negative; at right, the areas bounded by f on the interval $[a, d]$.

For the function pictured in the first graph of Figure 4.2.10, a left Riemann sum with 12 subintervals over $[a, d]$ is shown. The function is negative on the interval $b \leq x \leq c$, so at the four left endpoints that fall in $[b, c]$, the terms $f(x_i)\Delta x$ are negative. 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 $[b, c]$.

In the middle graph of Figure 4.2.10, we see that by increasing the number of rectangles the approximation of the area (or the opposite of the area) bounded by the curve appears to improve.

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 4.2.10, we may say that

$$L_{24} \approx A_1 - A_2 + A_3,$$

where L_{24} is the left Riemann sum using 24 subintervals shown in the middle graph. A_1 and A_3 are the areas of the regions where f is positive, and A_2 is the area where f is negative. We will 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 that if the function f represents the velocity of a moving object, the sum of the areas bounded by the curve tells us the total distance traveled over the relevant time interval, while the net signed area bounded by the curve computes the object’s change in position on the interval.

Example 4.2.11. 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 Δt as well as the locations of t_0, t_1, \dots, t_5 . In addition, pro-

vide a careful sketch of the function and the corresponding rectangles that are being used in the sum.

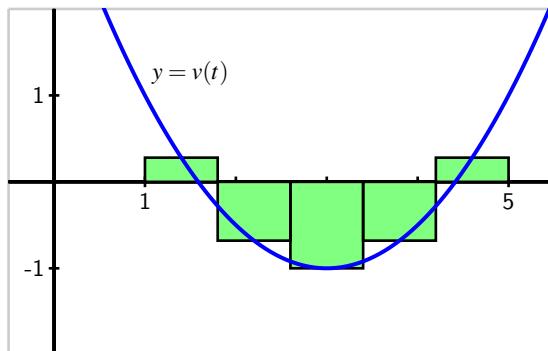
- Building on your work in (a), estimate the total change in position of the object on the interval $[1, 5]$.
- Building on your work in (a) and (b), estimate the total distance traveled by the object on $[1, 5]$.
- Use appropriate computing technology² to compute M_{10} and M_{20} . What exact value 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?

Hint.

- Note that $\Delta t = \frac{5-1}{5}$.
- Change in position is tied to the net signed area bounded by the velocity function.
- Think about how total distance is different from change in position when the velocity is sometimes negative.
- Besides the noted applet, computer algebra systems such as *Maple* and *Mathematica* offer this utility.

Answer.

a. $M_5 = -\frac{36}{25} = -1.44$



- The change in position is approximately -1.44 feet.
- $D \approx 2.336$.
- $-\frac{4}{3}$ is the object's total change in position on $[1, 5]$.

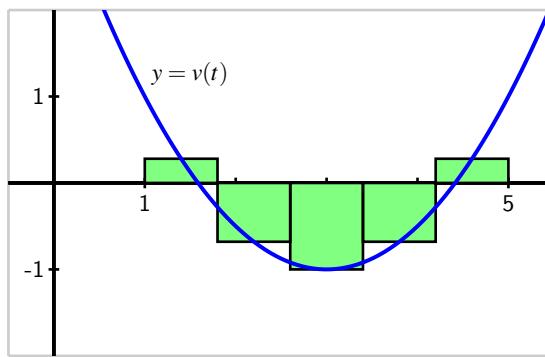
²For instance, consider the applet at <http://gvsu.edu/s/a9> and change the function and adjust the locations of the blue points that represent the interval endpoints a and b .

Solution.

- a. For this Riemann sum with five subintervals, $\Delta t = \frac{5-1}{5} = \frac{4}{5}$, so $t_0 = 1$, $t_1 = 1.8$, $t_2 = 2.6$, $t_3 = 3.4$, $t_4 = 4.2$ and $t_5 = 4$. It follows that

$$\begin{aligned} M_5 &= v(1.4)\Delta t + v(2.2)\Delta t + v(3.0)\Delta t + v(3.8)\Delta t + v(4.6) \\ &= \frac{7}{25} \cdot \frac{4}{5} - \frac{17}{25} \cdot \frac{4}{5} - 1 \cdot \frac{4}{5} - \frac{17}{25} \cdot \frac{4}{5} + \frac{7}{25} \cdot \frac{4}{5} \\ &= -\frac{36}{25} = -1.44 \end{aligned}$$

A sketch is shown below.



- b. Since the net signed area bounded by v on $[1, 5]$ represents the total change in position of the object on the interval $[1, 5]$, it follows that M_5 estimates the total change in position. Hence, the change in position is approximately -1.44 feet.
- c. To estimate the total distance traveled by the object on $[1, 5]$, we have to calculate the total area between the curve and the t -axis. Thus,

$$D \approx \frac{7}{25} \cdot \frac{4}{5} + \frac{17}{25} \cdot \frac{4}{5} + 1 \cdot \frac{4}{5} + \frac{17}{25} \cdot \frac{4}{5} + \frac{7}{25} \cdot \frac{4}{5} = \frac{292}{125} \approx 2.336.$$

- d. Using appropriate technology, $M_{10} = -1.36$ and $M_{20} = -1.34$. Further calculations suggest that $M_n \rightarrow -\frac{4}{3} = -1.\overline{33}$ as $n \rightarrow \infty$, and this number represents the object's total change in position on $[1, 5]$.

4.2.4 Summary

- A Riemann sum is simply a sum of products of the form $f(x_i^*)\Delta x$ 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, but we can also work with a more general 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. 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

$$L_n = f(x_0)\Delta x + f(x_1)\Delta x + \cdots + f(x_{n-1})\Delta x = \sum_{i=0}^{n-1} f(x_i)\Delta x,$$

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = \sum_{i=1}^n f(x_i)\Delta x,$$

$$M_n = f(\bar{x}_1)\Delta x + f(\bar{x}_2)\Delta x + \cdots + f(\bar{x}_n)\Delta x = \sum_{i=1}^n f(\bar{x}_i)\Delta x,$$

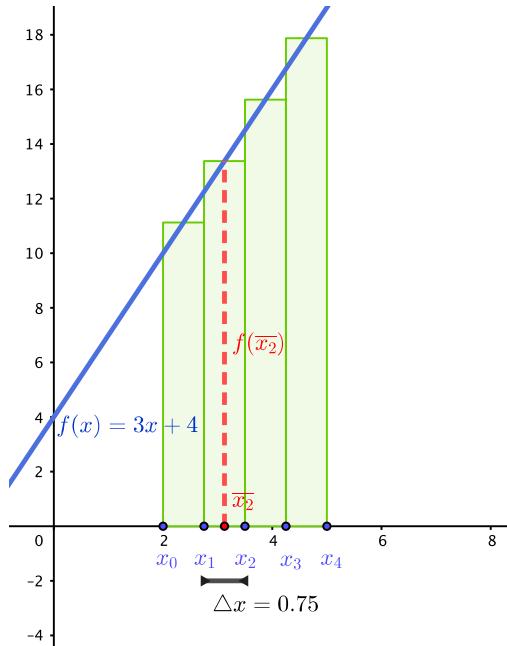
where $x_0 = a$, $x_i = a + i\Delta x$, and $x_n = b$, using $\Delta x = \frac{b-a}{n}$. For the midpoint sum, $\bar{x}_i = (x_{i-1} + x_i)/2$.

4.2.5 Exercises

1. (Evaluating Riemann sums for a quadratic function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Estimating distance traveled with a Riemann sum from data) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Writing basic Riemann sums) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. 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 Δx , as well as the locations of x_0, x_1, \dots, 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 $[a, b]$.

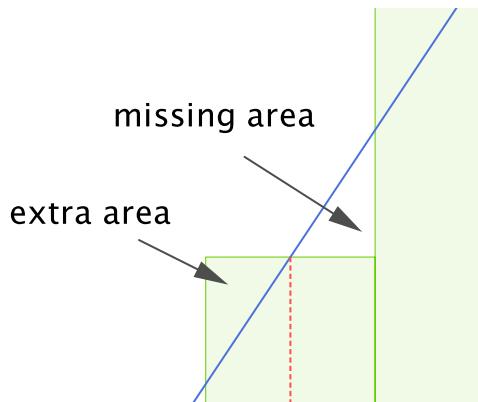
Answer.

- a. $M_4 = 43.5$.



b. $A = \frac{87}{2}$.

- c. The rectangles with heights that come from the midpoint have the same area as the trapezoids that are formed by the function values at the two endpoints of each subinterval.



M_n will give the exact area for any value of n . Neither L_n nor R_n will be exact for any n .

- d. For any linear function g of the form $g(x) = mx + b$ such that $g(x) \geq 0$ on the interval of interest.

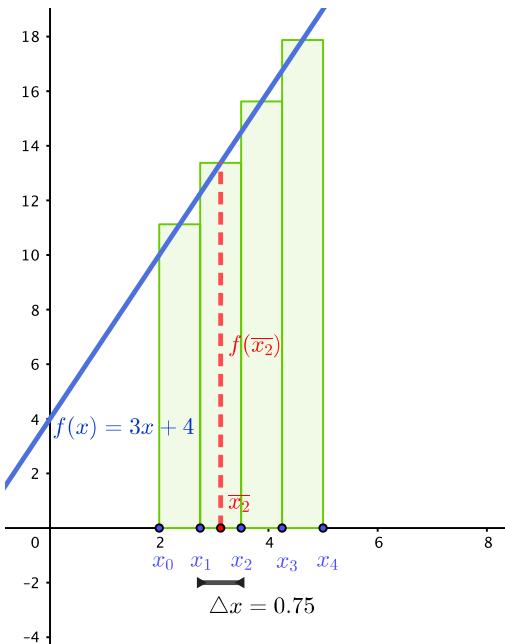
Solution.

- a. Since we are using 4 subintervals on $[2, 5]$, it follows that $\Delta x = \frac{5-2}{4} = 0.75$. Thus, $x_0 = 2$, $x_1 = 2.75$, $x_2 = 3.5$, $x_3 = 4.25$, and $x_4 = 5$. The midpoints of the respective subintervals are thus $\bar{x}_1 = \frac{2+2.75}{2} = 2.375$, $\bar{x}_2 = 3.125$, $\bar{x}_3 = 3.875$, and $\bar{x}_4 = 4.625$. Thus,

$$M_4 = f(\bar{x}_1) \cdot \Delta x + f(\bar{x}_2) \cdot \Delta x + f(\bar{x}_3) \cdot \Delta x + f(\bar{x}_4) \cdot \Delta x$$

$$\begin{aligned}
 &= (f(2.375) + f(3.125) + f(3.875) + f(4.625)) \cdot (0.75) \\
 &= (11.125 + 13.375 + 15.625 + 17.875) \cdot (0.75) \\
 &= 58 \cdot 0.75 \\
 &= 43.5
 \end{aligned}$$

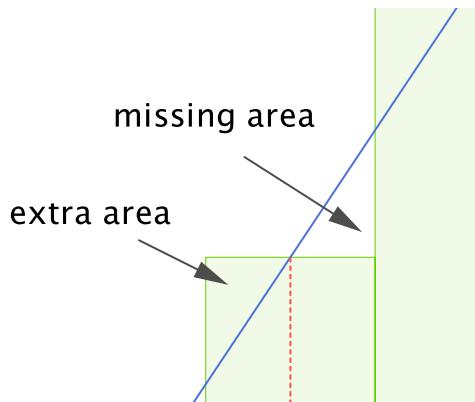
An overall sketch of the key parts of M_4 is seen in the figure below.



- b. Observe that $f(2) = 10$ and $f(5) = 19$. The region bounded by the function, the x -axis, and the endpoints is a trapezoid. Recalling that the area of a trapezoid is the average of the bases times the height, we see that the bounded region has area

$$A = \frac{1}{2}(10 + 19)(5 - 2) = \frac{87}{2}.$$

- c. The reason that M_4 gives the exact area between the line and the x -axis is the rectangles with heights that come from the midpoint have the same area as the trapezoids that are formed by the function values at the two endpoints of each subinterval. We can also see this by observing that the “extra area” bounded by the midpoint rectangle exactly matches the “missing area” that is absent from the midpoint rectangle, as seen in the following figure. Since these two quantities balance out, the areas of the midpoint rectangles give us the exact area under the curve.



Indeed, the value of M_n will give the exact area under g on $[2, 5]$ for any value of n . But neither L_n nor R_n will be exact: L_n will underestimate the true area since every rectangle will lie completely under the function g , whereas R_n will overestimate because every rectangle lies above the function.

- d. For any linear function g of the form $g(x) = mx + b$ such that $g(x) \geq 0$ on the interval of interest, the midpoint rule will give the exact value of the area bounded by the function and the x -axis. This is true because the area of the trapezoid formed by the two endpoints on a given subinterval will be the same as the area of the rectangle whose height is given by the function at the midpoint of the subinterval.

5. Let S be the sum given by

$$S = ((1.4)^2 + 1) \cdot 0.4 + ((1.8)^2 + 1) \cdot 0.4 + ((2.2)^2 + 1) \cdot 0.4 + ((2.6)^2 + 1) \cdot 0.4 + ((3.0)^2 + 1) \cdot 0.4.$$

- Assume that S is a right Riemann sum. For what function f and what interval $[a, b]$ is S this function's Riemann sum? Why?
- How does your answer to (a) change if S is a left Riemann sum? a middle Riemann sum?
- Suppose that S really is a right Riemann sum. What is geometric quantity does S approximate?
- 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 .

Answer.

- $f(x) = x^2 + 1$ on the interval $[1, 3]$.
- If S is a left Riemann sum, $f(x) = x^2 + 1$ on the interval $[1.4, 3.4]$. If S is a middle Riemann sum, $f(x) = x^2 + 1$ on the interval $[1.2, 3.2]$.
- The area under $f(x) = x^2 + 1$ on $[1, 3]$.
- $R_{10} = \sum_{i=1}^{10} ((1 + 0.2i)^2 + 1) \cdot 0.2$.

Solution.

- The function is being evaluated at $x_1 = 1.4$, $x_2 = 1.8$, $x_3 = 2.2$, $x_4 = 2.6$, and $x_5 = 3.0$. Thus, since we know this is a right Riemann sum, we are using 5 subintervals with $\Delta x = 0.4$, and $x_0 = 1.0$. From the pattern, we observe that the function must be $f(x) = x^2 + 1$ on the interval $[1, 3]$.

- b. If S is a left Riemann sum instead, then the x_i 's shift left: $x_0 = 1.4, x_1 = 1.8, \dots, x_5 = 3.4$. In this setting, the function is identical, but we are considering it on the interval $[1.4, 3.4]$.

If S is a middle Riemann sum, then (for example) 1.4 is the midpoint of an interval $\Delta x = 0.4$ units long, so $x_0 = 1.2$ and $x_1 = 1.6$. Continuing that trend, we'd be considering the function $f(x) = x^2 + 1$ on the interval $[1.2, 3.2]$.

- c. S is an approximation of the area that lies under $f(x) = x^2 + 1$ on $[1, 3]$.
- d. To use twice as many subintervals, we move to $\Delta x = 0.2$ and thus $x_1 = 1.2, x_2 = 1.4$, and so on, with the general right endpoint being given by $x_i = 1 + 0.2 \cdot i$. We'll also have $n = 10$, so

$$R_{10} = \sum_{i=1}^{10} ((1 + 0.2i)^2 + 1) \cdot 0.2.$$

6. A car traveling along a straight road is braking and its velocity is measured at several different points in time, as given in the following table.

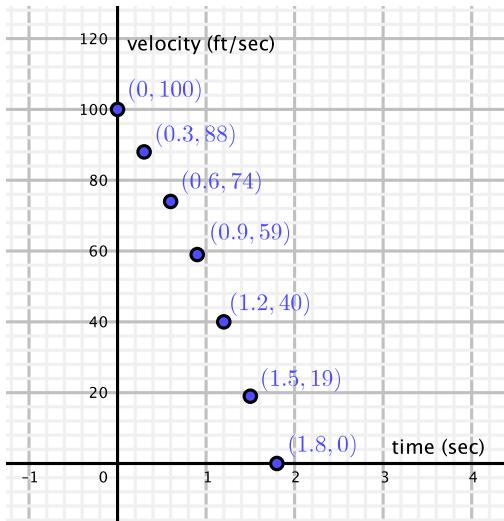
seconds, t	0	0.3	0.6	0.9	1.2	1.5	1.8
Velocity in ft/sec, $v(t)$	100	88	74	59	40	19	0

Table 4.2.12: Data for the braking car.

- 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?

Answer.

- a.



b. $M_3 = 99.6$ feet.

c. $L_6 = 114$,

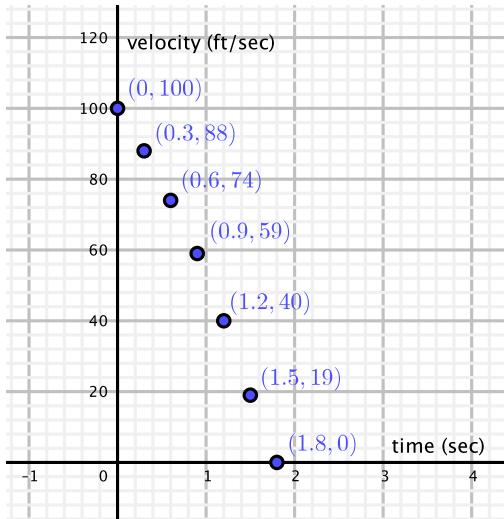
$$R_6 = 84,$$

and $\frac{1}{2}(L_6 + R_6) = 99$.

d. 114 feet.

Solution.

a. A plot of the given data is shown in the figure below.



b. For M_3 with the given data, the midpoints are $\bar{x}_1 = 0.3$, $\bar{x}_2 = 0.9$, and $\bar{x}_3 = 1.5$, and $\Delta t = 0.6$, so

$$M_3 = f(0.3) \cdot 0.6 + f(0.9) \cdot 0.6 + f(1.5) \cdot 0.6 = 166 \cdot 0.6 = 99.6.$$

Therefore, the car travels approximately 99.6 feet over the 1.8 second interval it brakes.

- c. Using the standard formulas,

$$L_6 = f(0) \cdot 0.3 + f(0.3) \cdot 0.3 + \cdots + f(1.5) \cdot 0.3 = 380 \cdot 0.3 = 114$$

and

$$R_6 = f(0.3) \cdot 0.3 + f(0.6) \cdot 0.3 + \cdots + f(1.8) \cdot 0.3 = 280 \cdot 0.3 = 84.$$

Thus,

$$\frac{1}{2}(L_6 + R_6) = \frac{114 + 84}{2} = 99.$$

Each of these three quantities is an estimate of the total distance traveled: L_6 is an overestimate and R_6 is an underestimate.

- d. The maximum possible distance the car traveled before it stopped is 114 feet. This is because L_n is always an overestimate when the function is decreasing on the interval.
7. 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 in Figure 4.2.13.
- Use the graph to estimate the value of M_4 on the interval $[0, 4]$.
 - What is the meaning of M_4 in terms of the pollution discharged by the plant?
 - Suppose that $r(t) = 0.5e^{0.5t}$. Use this formula for r to compute L_5 on $[0, 4]$.
 - 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.

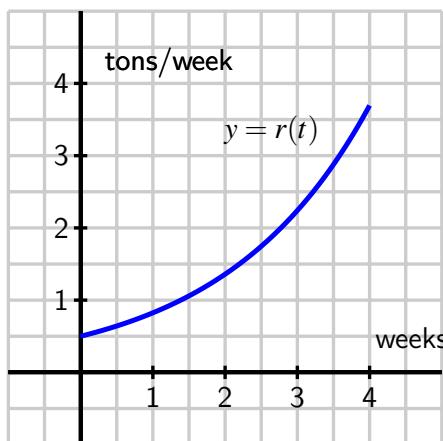
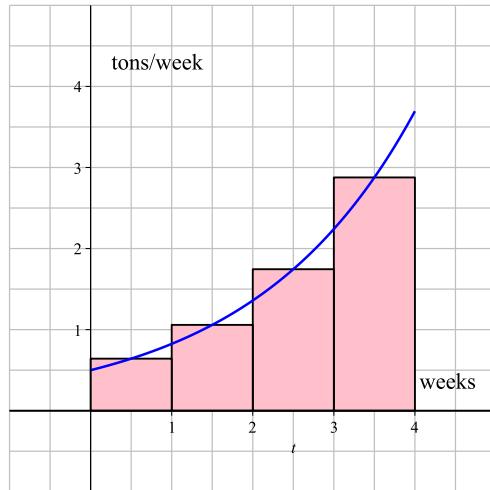


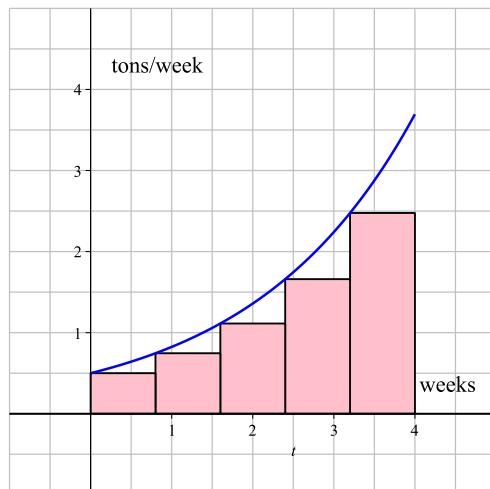
Figure 4.2.13: The rate, $r(t)$, of pollution in tons per week.

Answer.

- a. $M_4 \approx 6.4$.



- b. The total tonnage of pollution escaping the scrubbing process in the time interval $[0, 4]$ weeks.
 c. $L_5 \approx 5.19620599$.

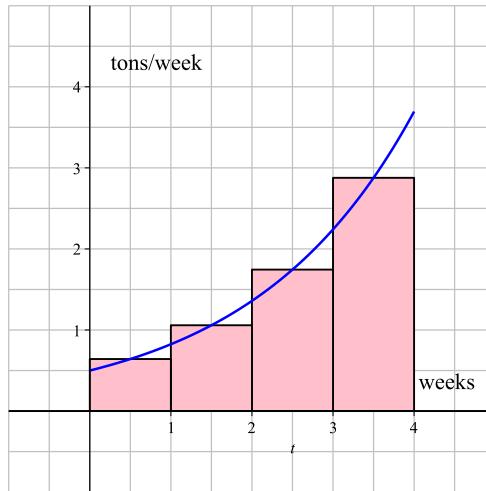


- d. 6.4 tons.

Solution.

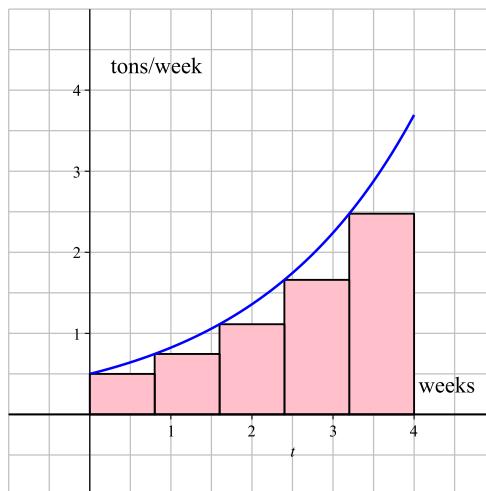
- a. We estimate the value of r at the midpoint of each subinterval and then multiply these values by the width (1) of each subinterval to estimate the areas of the rectangles shown in the following figure. We use the following approximations: $r(0.5) \approx 0.6$, $r(1.5) \approx 1.1$, $r(2.5) \approx 1.8$, $r(3.5) \approx 2.9$, so

$$M_4 \approx (0.6)(1) + (1.1)(1) + (1.8)(1) + (2.9)(1) = 6.4.$$



- b. Since $r(t)$ is measured in tons/week and t in weeks, the units of $r(t)\Delta t$ are tons. So M_4 approximates the total tonnage of pollution escaping the scrubbing process in the time interval $[0, 4]$ weeks.
- c. Now we are using 5 subintervals, so $\Delta t = \frac{4}{5} = 0.8$. We evaluate r at the left endpoint of each subinterval and multiply by Δt to determine the areas of the rectangles shown in the following figure. So

$$\begin{aligned}
 L_5 &= \left(0.5e^{(0.5)(0)}\right)(0.8) + \left(0.5e^{(0.5)(0.8)}\right)(0.8) + \left(0.5e^{(0.5)(1.6)}\right)(0.8) \\
 &\quad + \left(0.5e^{(0.5)(2.4)}\right)(0.8) + \left(0.5e^{(0.5)(3.2)}\right)(0.8) \\
 &\approx 5.19620599.
 \end{aligned}$$



- d. Given that r is an increasing function, a right sum will always give us an overestimate of the total amount of pollution (in tons) on the interval $[0, 4]$. We use computational

technology to compute a few right sums: $R_{10} \approx 7.049244376$, $R_{50} \approx 6.517689073$, $R_{100} \approx 6.453159616$, $R_{500} \approx 6.401842731$. These sums don't seem to change in the tens place as we make n sufficiently large, so an upper bound on the total amount of pollution on the interval $[0, 4]$ that is accurate to within one ton is 6.4 tons.

4.3 The Definite Integral

Motivating 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?

In Figure 4.3.1, we see evidence that increasing the number of rectangles in a Riemann sum improves the accuracy of the approximation of the net signed area bounded by the given function.

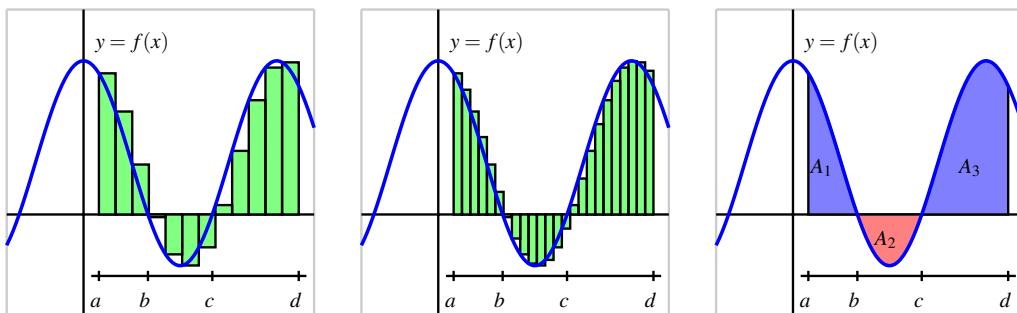


Figure 4.3.1: At left and center, two left Riemann sums for a function f that is sometimes negative; at right, the exact areas bounded by f on the interval $[a, d]$.

We therefore explore the natural idea of allowing the number of rectangles to increase without bound. In an effort to compute the exact net signed area we also consider the differences among left, right, and middle Riemann sums and the different results they generate as the value of n increases. We begin with functions that are exclusively positive on the interval under consideration.

Example 4.3.2. Consider the applet found at <http://gvsu.edu/s/a9>¹. There, you will initially see the situation shown in Figure 4.3.3.

¹Marc Renault, Shippensburg University, Geogebra Applets for Calculus, <http://gvsu.edu/s/5p>.

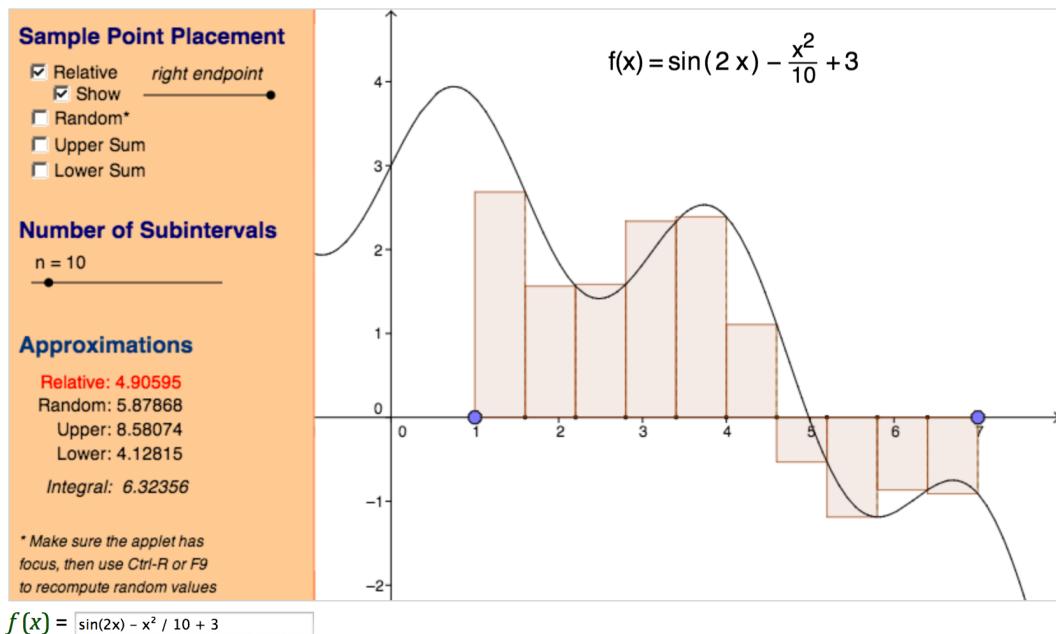


Figure 4.3.3: A right Riemann sum with 10 subintervals for the function $f(x) = \sin(2x) - \frac{x^2}{10} + 3$ on the interval $[1, 7]$. The value of the sum is $R_{10} = 4.90595$.

Note that the value of the chosen Riemann sum is displayed next to the word “relative,” and that you can change the type of Riemann sum being computed by dragging the point on the slider bar below the phrase “sample point placement.”

Explore to see how you can change the window in which the function is viewed, as well as the function itself. You can set the minimum and maximum values of x by clicking and dragging on the blue points that set the endpoints; you can change the function by typing a new formula in the “ $f(x)$ ” window at the bottom; and you can adjust the overall window by “panning and zooming” by using the Shift key and the scrolling feature of your mouse. More information on how to pan and zoom can be found at <http://gvsu.edu/s/F1>.

Work accordingly to adjust the applet so that it uses a left Riemann sum with $n = 5$ subintervals for the function is $f(x) = 2x + 1$. You should see the updated figure shown in Figure 4.3.4. Then, answer the following questions.

- Update the applet (and view window, as needed) 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 $n = 5$, $n = 25$, and $n = 100$. What appears to be the exact area bounded by $f(x) = 2x + 1$ and the x -axis on $[1, 4]$?
- Use basic geometry to determine the exact area bounded by $f(x) = 2x + 1$ and the x -axis on $[1, 4]$.
- 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 need to enter “ $x^2 + 1$ ” for the function formula). Use the applet to compute L_n, M_n, R_n for $n = 5, n = 25$, and $n = 100$. 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)?

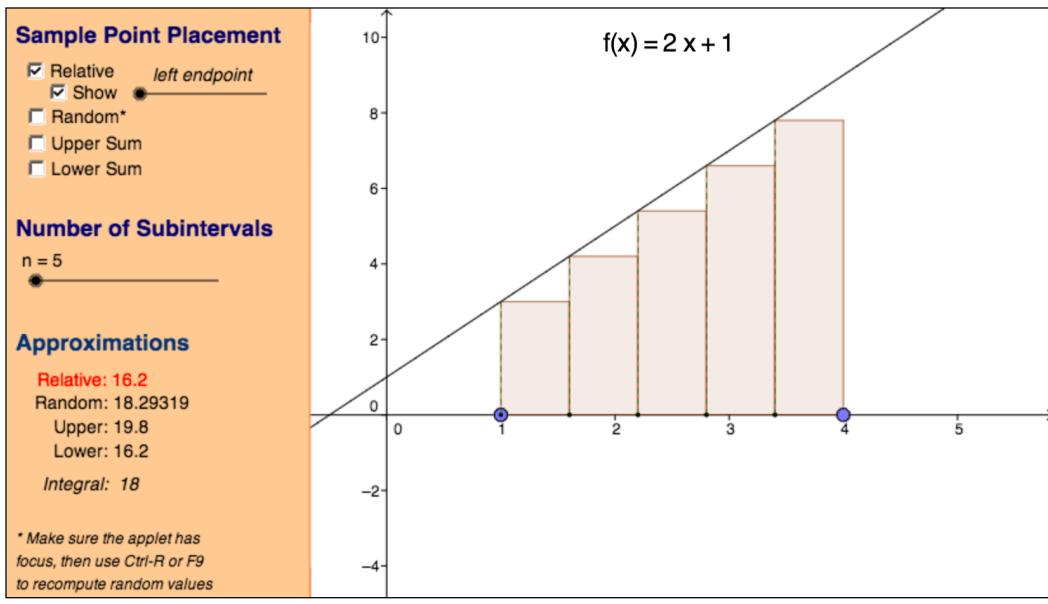


Figure 4.3.4: A left Riemann sum with 5 subintervals for the function $f(x) = 2x + 1$ on the interval $[1, 4]$. The value of the sum is $L_5 = 16.2$.

4.3.1 The definition of the definite integral

In Example 4.3.2, 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 also for a Riemann sum using any point x_{i+1}^* in the interval $[x_i, x_{i+1}]$. Thus, as we let $n \rightarrow \infty$, it doesn't really matter where we choose to evaluate the function within a given subinterval, because

$$\lim_{n \rightarrow \infty} L_n = \lim_{n \rightarrow \infty} R_n = \lim_{n \rightarrow \infty} M_n = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \Delta x.$$

That these limits always exist (and share the same value) when f is continuous² allows us to make the following definition.

²It turns out that a function need not be continuous in order to have a definite integral. For our purposes, we

Definition 4.3.5. The **definite integral** of a continuous function f on the interval $[a, b]$, denoted $\int_a^b f(x) dx$, is the real number given by

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \Delta x,$$

where $\Delta x = \frac{b-a}{n}$, $x_i = a + i\Delta x$ (for $i = 0, \dots, n$), and x_i^* satisfies $x_{i-1} \leq x_i^* \leq x_i$ (for $i = 1, \dots, n$).

We call the symbol \int the *integral sign*, the values a and b the *limits 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 there are several different interpretations 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 $[a, b]$.

For example, if f is the function pictured in Figure 4.3.6, 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

$$\int_a^b f(x) dx = A_1, \quad \int_b^c f(x) dx = -A_2,$$

$$\int_c^d f(x) dx = A_3,$$

and $\int_a^d f(x) dx = A_1 - A_2 + A_3$.

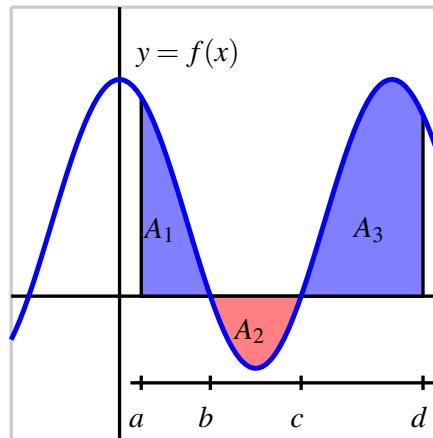


Figure 4.3.6: A continuous function f on the interval $[a, d]$.

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.

We can also use definite integrals to express the change in position and the distance traveled by a moving object. If v is a velocity function on an interval $[a, b]$, then the change in position of the object, $s(b) - s(a)$, 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. If the velocity is sometimes negative on $[a, b]$, we can use definite integrals to find the areas bounded by the function on each interval where v does not change sign, and the sum of these areas will tell us the distance the object traveled.

To compute the value of a definite integral from 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, and does not offer much additional insight into the meaning or interpretation of the definite integral. Instead, in Section 4.4, we will learn the Fundamental Theorem of Calculus, which provides a shortcut 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.

For now, our goal is to understand the meaning and properties of the definite integral, rather than to compute its value. To do this, we will rely on the net signed area interpretation of the definite integral. So we will use as examples curves that produce regions whose areas we can compute exactly through area formulas. We can thus compute the exact value of the corresponding integral.

For instance, if we wish to evaluate the definite integral $\int_1^4 (2x + 1) dx$, we observe that the region bounded by this function and the x -axis is the trapezoid shown in Figure 4.3.7. By the formula for the area of a trapezoid, $A = \frac{1}{2}(3 + 9) \cdot 3 = 18$, so

$$\int_1^4 (2x + 1) dx = 18.$$

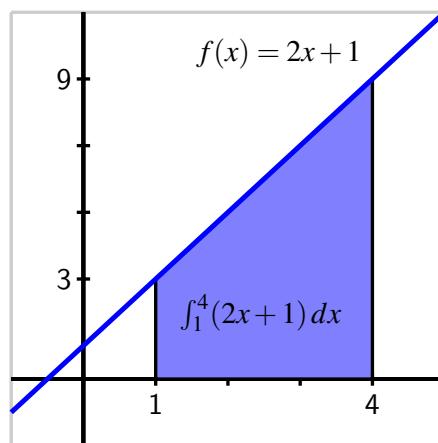


Figure 4.3.7: The area bounded by $f(x) = 2x + 1$ and the x -axis on the interval $[1, 4]$.

Example 4.3.8. Use known geometric formulas and the net signed area interpretation of the definite integral to evaluate each of the definite integrals below.

a. $\int_0^1 3x \, dx$

b. $\int_{-1}^4 (2 - 2x) \, dx$

c. $\int_{-1}^1 \sqrt{1 - x^2} \, dx$

d. $\int_{-3}^4 g(x) \, dx$, where g is the function pictured in Figure 4.3.9. Assume that each portion of g is either part of a line or part of a circle.

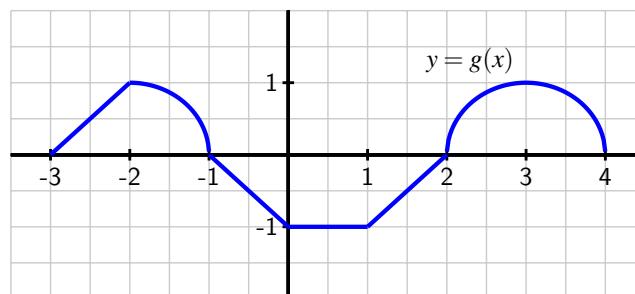


Figure 4.3.9: A function g that is piecewise defined; each piece of the function is part of a circle or part of a line.

Hint.

- Sketch the region bounded by $y = 3x$ and the x -axis on $[0, 1]$.
- Sketch the region bounded by $y = 2 - 2x$ and the x -axis on $[-1, 4]$.
- Observe that $y = \sqrt{1 - x^2}$ is the top half the circle whose equation is $x^2 + y^2 = 1$.
- Use known formulas for the area of a triangle, square, or circle appropriately.

Answer.

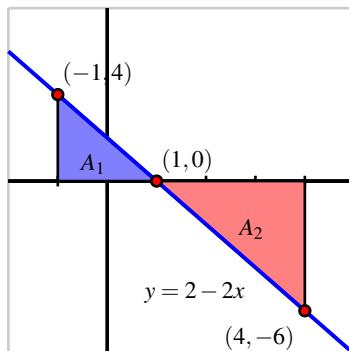
- $\int_0^1 3x \, dx = \frac{3}{2}$.
- $\int_{-1}^4 (2 - 2x) \, dx = -5$.
- $\int_{-1}^1 \sqrt{1 - x^2} \, dx = \frac{\pi}{2}$.
- $\int_{-3}^4 g(x) \, dx = \frac{3\pi}{4} - \frac{3}{2}$.

Solution.

- Because $y = 3x$ and the x -axis bound a triangle with base of length 1 and height 3 on the interval $[0, 1]$, it follows that

$$\int_0^1 3x \, dx = \frac{1}{2} \cdot 1 \cdot 3 = \frac{3}{2}.$$

- For $\int_{-1}^4 (2 - 2x) \, dx$, we first sketch the region bounded by the function, as shown below.



The line creates two triangles, one with area $A_1 = \frac{1}{2} \cdot 2 \cdot 4 = 4$ and the other with area $A_2 = \frac{1}{2} \cdot 3 \cdot 6 = 9$. Since the latter area corresponds to a region below the x -axis, we associate a negative sign to it, and hence find that

$$\int_{-1}^4 (2 - 2x) \, dx = A_1 - A_2 = 4 - 9 = -5.$$

- c. For $\int_{-1}^1 \sqrt{1-x^2} dx$, we simply observe that this function is the top half of a circle of radius 1, and thus the bounded region is a semicircle of radius 1, thus having an area of $\frac{\pi}{2}$. Therefore,

$$\int_{-1}^1 \sqrt{1-x^2} dx = \frac{\pi}{2}.$$

- d. Finally, for $\int_{-3}^4 g(x) dx$, where g is the function pictured in Figure 4.3.9, we consider the function on seven consecutive subintervals of length 1. Observe that on $[-3, -2]$, the bounded area is $\frac{1}{2}$. On $[-2, -1]$, the area is $\frac{1}{4}\pi$. Similarly, on the next five subintervals of length 1, the areas bounded are respectively $\frac{1}{2}$, 1, $\frac{1}{2}$, $\frac{1}{4}\pi$, and $\frac{1}{4}\pi$. Thus, the value of the integral is

$$\int_{-3}^4 g(x) dx = \frac{1}{2} + \frac{\pi}{4} - \frac{1}{2} - 1 - \frac{1}{2} + \frac{\pi}{4} + \frac{\pi}{4} = \frac{3\pi}{4} - \frac{3}{2},$$

which is approximately 0.8562.

4.3.2 Some properties of the definite integral

Regarding the definite integral of a function f over an interval $[a, b]$ as the net signed area bounded by f and the x -axis, we discover several standard properties of the definite integral. It is helpful to remember that the definite integral is defined in terms of Riemann sums, which consist of the areas of rectangles.

For any real number a and the definite integral $\int_a^a f(x) dx$ it is evident that no area is enclosed, because the interval begins and ends with the same point. Hence,

If f is a continuous function and a is a real number, then $\int_a^a f(x) dx = 0$.

Next, we consider the result of subdividing the interval of integration. In Figure 4.3.10, we see that

$$\int_a^b f(x) dx = A_1, \quad \int_b^c f(x) dx = A_2,$$

and $\int_a^c f(x) dx = A_1 + A_2,$

which illustrates the following general rule.

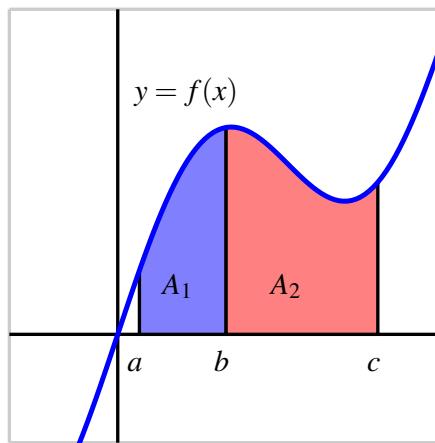


Figure 4.3.10: The area bounded by $y = f(x)$ on the interval $[a, c]$.

If f is a continuous function and a , b , and c are real numbers, then

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx.$$

While this rule is easy to see if $a < b < c$, it in fact holds in general for any values of a , b , and c . Another property of the definite integral states that if we reverse the order of the limits of integration, we change the sign of the integral's value.

If f is a continuous function and a and b are real numbers, then

$$\int_b^a f(x) dx = - \int_a^b f(x) dx.$$

This result makes sense because if we integrate from a to b , then in the defining Riemann sum we set $\Delta x = \frac{b-a}{n}$, while if we integrate from b to a , we have $\Delta x = \frac{a-b}{n} = -\frac{b-a}{n}$, and this is the only change in the sum used to define the integral.

There are two additional useful properties of the definite integral. When we worked with derivative rules in Chapter 2, we formulated the Constant Multiple Rule and the Sum Rule. Recall that the Constant Multiple Rule says that if f is a differentiable function and k is a constant, then

$$\frac{d}{dx} [kf(x)] = kf'(x),$$

and the Sum Rule says that if f and g are differentiable functions, then

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

These rules are useful because they allow to deal individually with the simplest parts of certain functions by taking advantage of addition and multiplying by a constant. In other words, the process of taking the derivative respects addition and multiplying by constants in the simplest possible way.

It turns out that similar rules hold for the definite integral. First, let's consider the functions pictured in Figure 4.3.11.

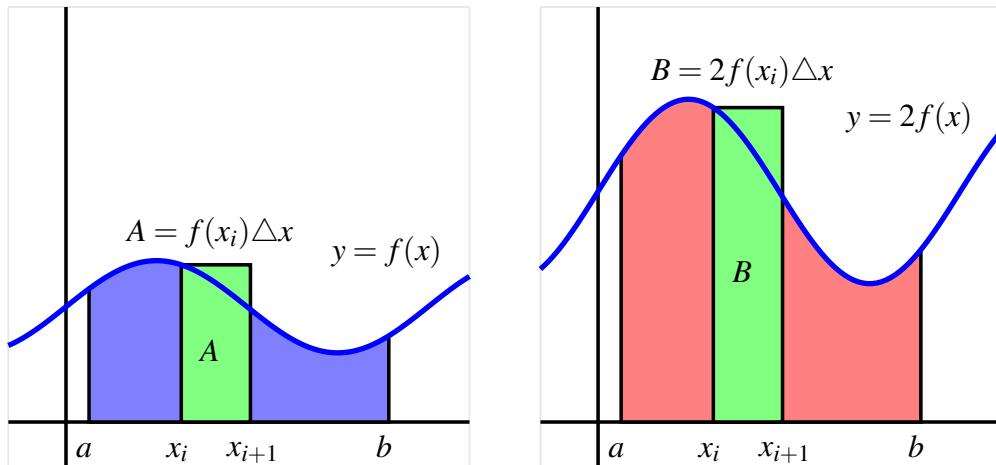


Figure 4.3.11: The areas bounded by $y = f(x)$ and $y = 2f(x)$ on $[a, b]$.

Because multiplying the function by 2 doubles its height at every x -value, we see that the height of each rectangle in a left Riemann sum is doubled, $f(x_i)$ for the original function, versus $2f(x_i)$ in the doubled function. For the areas A and B , it follows $B = 2A$. As this is true regardless of the value of n or 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 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 , the following general principle holds.

Constant Multiple Rule

If f is a continuous function and k is any real number, then

$$\int_a^b k \cdot f(x) dx = k \int_a^b f(x) dx.$$

We see a similar situation with the sum of two functions f and g .

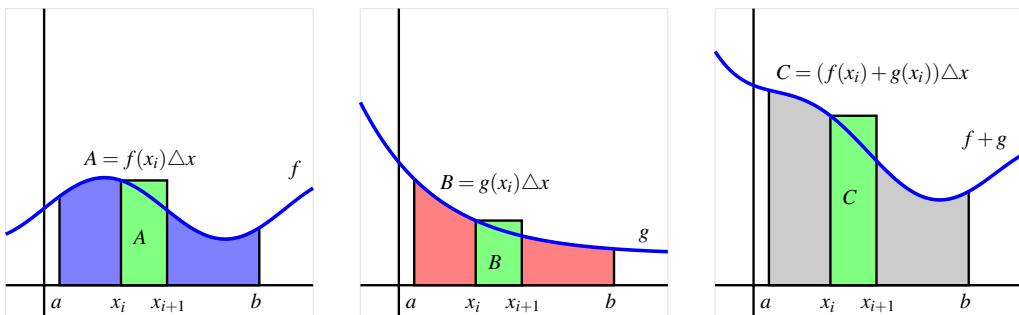


Figure 4.3.12: The areas bounded by $y = f(x)$ and $y = g(x)$ on $[a, b]$, as well as the area bounded by $y = f(x) + g(x)$.

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 $(f + g)(x_i) = f(x_i) + g(x_i)$. Hence, for the pictured rectangles with areas A , B , and C , it follows that $C = A + B$. 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. In terms of definite integrals, we have the following general rule.

Sum Rule

If f and g are continuous functions, then

$$\int_a^b [f(x) + g(x)] dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$$

The Constant Multiple and Sum Rules can be combined to say that for any continuous functions f and g and any constants c and k ,

$$\int_a^b [cf(x) \pm kg(x)] dx = c \int_a^b f(x) dx \pm k \int_a^b g(x) dx.$$

Example 4.3.13. 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. $\int_5^2 f(x) dx$ | d. $\int_2^5 (3x^2 - 4x^3) dx$ |
| b. $\int_0^5 g(x) dx$ | |
| c. $\int_0^5 (f(x) + g(x)) dx$ | e. $\int_5^0 (2x^3 - 7g(x)) dx$ |

Hint.

- a. Note that the value of $\int_2^5 f(x) dx$ is given.
- b. Use the values of $\int_0^2 g(x) dx$ and $\int_2^5 g(x) dx$.
- c. First find $\int_0^5 f(x) dx$ and $\int_0^5 g(x) dx$.
- d. Use the sum and constant multiple rules.
- e. First write $\int_5^0 (2x^3 - 7g(x)) dx = -\int_0^5 (2x^3 - 7g(x)) dx$.

Answer.

- a. $\int_5^2 f(x) dx = -2$.
- b. $\int_0^5 g(x) dx = 3$.
- c. $\int_0^5 (f(x) + g(x)) dx = 2$.
- d. $\int_2^5 (3x^2 - 4x^3) dx = -492$.
- e. $\int_5^0 (2x^3 - 7g(x)) dx = -\frac{187}{3}$.

Solution.

- a. Note that the value of $\int_2^5 f(x) dx$ is given, and thus

$$\int_5^2 f(x) dx = - \int_2^5 f(x) dx = -2.$$

- b. Since $\int_0^2 g(x) dx = 4$ and $\int_2^5 g(x) dx = -1$, we have

$$\int_0^5 g(x) dx = \int_0^2 g(x) dx + \int_2^5 g(x) dx = 4 + (-1) = 3.$$

- c. First, using work from and similar to that in (c), we find $\int_0^5 f(x) dx = -3 + 2 = -1$ and $\int_0^5 g(x) dx = 3$, thus by the sum rule,

$$\int_0^5 (f(x) + g(x)) dx = \int_0^5 f(x) dx + \int_0^5 g(x) dx = -1 + 3 = 2.$$

d. By the sum and constant multiple rules,

$$\int_2^5 (3x^2 - 4x^3) dx = 3 \int_2^5 x^2 dx - 4 \int_2^5 x^3 dx = 3 \cdot \frac{117}{3} - 4 \cdot \frac{609}{4} = 117 - 609 = -492.$$

e. First, we write $\int_5^0 (2x^3 - 7g(x)) dx = - \int_0^5 (2x^3 - 7g(x)) dx$. Then, using the sum and constant multiple rules, it follows

$$\begin{aligned} \int_5^0 (2x^3 - 7g(x)) dx &= - \int_0^5 (2x^3 - 7g(x)) dx \\ &= - \left(2 \int_0^5 x^3 dx - 7 \int_0^5 g(x) dx \right) \\ &= - 2 \left(\frac{8}{3} + \frac{117}{3} \right) + 7(4 + (-1)) \\ &= - \frac{250}{3} + 21 \\ &= - \frac{187}{3}. \end{aligned}$$

4.3.3 How the definite integral is connected to a function's average value

One of the most valuable applications of the definite integral is that it provides a way to discuss the average value of a function, even for a function that takes on infinitely many values. Recall that if we wish to take the average of n numbers y_1, y_2, \dots, y_n , we compute

$$AVG = \frac{y_1 + y_2 + \dots + y_n}{n}.$$

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 similar to averaging the output values of a function. Consider, for instance, the right Riemann sum R_n of a function f , which is given by

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \dots + f(x_n)\Delta x = (f(x_1) + f(x_2) + \dots + f(x_n))\Delta x.$$

Since $\Delta x = \frac{b-a}{n}$, we can thus write

$$\begin{aligned} R_n &= (f(x_1) + f(x_2) + \dots + f(x_n)) \cdot \frac{b-a}{n} \\ &= (b-a) \frac{f(x_1) + f(x_2) + \dots + f(x_n)}{n}. \end{aligned} \tag{4.3.1}$$

We see that the right Riemann sum with n subintervals is just 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, the larger the value of n we use, the more accurate our average will be. Indeed, we will define the average value of f on $[a, b]$ to be

$$f_{AVG[a,b]} = \lim_{n \rightarrow \infty} \frac{f(x_1) + f(x_2) + \dots + f(x_n)}{n}.$$

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, $\lim_{n \rightarrow \infty} R_n = \int_a^b f(x) dx$, and thus taking the limit as $n \rightarrow \infty$ in Equation (4.3.1), we have that

$$\int_a^b f(x) dx = (b - a) \cdot f_{\text{AVG}}[a, b]. \quad (4.3.2)$$

Solving Equation (4.3.2) 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

$$f_{\text{AVG}}[a, b] = \frac{1}{b - a} \cdot \int_a^b f(x) dx.$$

Equation (4.3.2) 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, when the function f is nonnegative on $[a, b]$, Equation (4.3.2) has a natural visual interpretation.

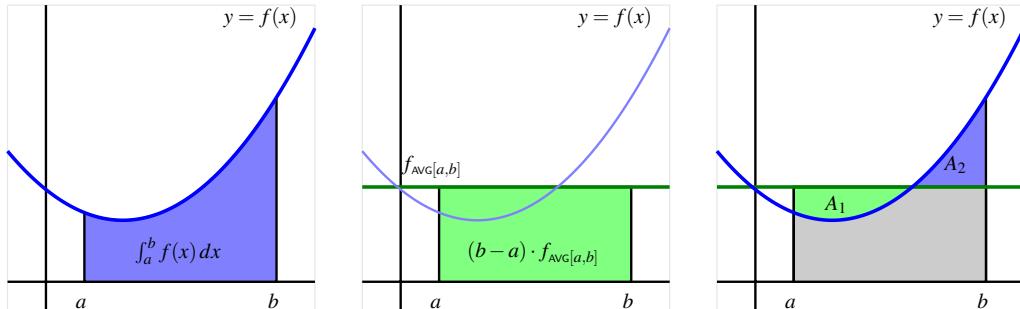


Figure 4.3.14: A function $y = f(x)$, the area it bounds, and its average value on $[a, b]$.

Consider Figure 4.3.14, where we see at left the shaded region whose area is $\int_a^b f(x) dx$, at center the shaded rectangle whose dimensions are $(b - a)$ by $f_{\text{AVG}}[a, b]$, and at right these two figures superimposed. 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 $\int_a^b f(x) dx$. The area of the blue region in the left figure is the same as the area of the green rectangle in the center figure. We can also observe that the areas A_1 and A_2 in the rightmost figure appear to be equal. Thus, knowing the average value of a function enables us to construct a rectangle whose area is the same as the value of the definite integral of the function on the interval. The java applet³ 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 4.3.14.

³David Austin, <http://gvsu.edu/s/5r>.

Example 4.3.15. Suppose that $v(t) = \sqrt{4 - (t - 2)^2}$ tells us the instantaneous velocity of a moving object on the interval $0 \leq t \leq 4$, where t is measured in minutes and v is measured in meters per minute.

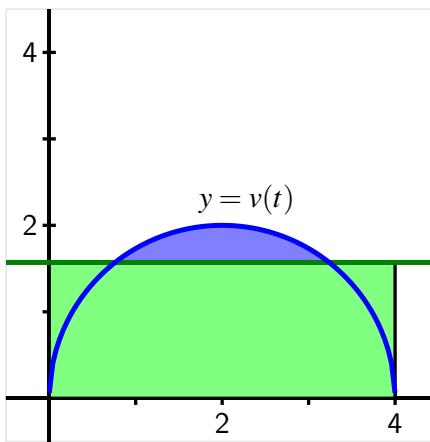
- Sketch an accurate graph of $y = v(t)$. What kind of curve is $y = \sqrt{4 - (t - 2)^2}$?
- Evaluate $\int_0^4 v(t) dt$ exactly.
- In terms of the physical problem of the moving object with velocity $v(t)$, what is the meaning of $\int_0^4 v(t) dt$? Include units in your answer.
- Determine the exact average value of $v(t)$ on $[0, 4]$. Include units in your answer.
- 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?
- How can you use the average value you found in (d) to compute the total distance traveled by the moving object over $[0, 4]$?

Hint.

- Note that $y = \sqrt{4 - (t - 2)^2}$ is part of the curve given by $(t - 2)^2 + y^2 = 4$.
- What familiar shape is generated by the curve $y = v(t)$?
- Recall the meaning of the area bounded by a nonnegative velocity function on a given interval.
- From the meaning of the average value of a function, we know $v_{\text{AVG}}[a, b] = \frac{1}{b-a} \int_a^b v(t) dt$.
- Consider a key recent figure in the text.
- Distance equals average rate times . . .

Answer.

- $y = v(t) = \sqrt{4 - (t - 2)^2}$ is the top half of the circle $(t - 2)^2 + y^2 = 4$, which has radius 2 and is centered at $(2, 0)$.
- $\int_0^4 v(t) dt = 2\pi$.
- The object moved 2π meters in 4 minutes.
- $v_{\text{AVG}}[0, 4] = \frac{\pi}{2}$, meters per minute.,
- e.



The height of the rectangle is the average value of v , $v_{\text{AVG}}[0, 4] = \frac{\pi}{2} \approx 1.57$.

- f. $D = 2\pi$.

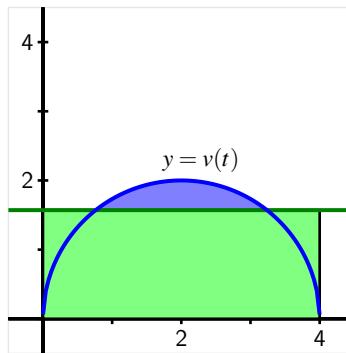
Solution.

- The curve $y = v(t) = \sqrt{4 - (t - 2)^2}$ is the top half of the circle $(t - 2)^2 + y^2 = 4$, which has radius 2 and is centered at $(2, 0)$.
- Thus, the value of $\int_0^4 v(t) dt$ is the area of a semicircle of radius 2, which is $\frac{1}{2}\pi(2)^2 = 2\pi$.
- Because the velocity $v(t)$ is always nonnegative in this problem, the meaning of $\int_0^4 v(t) dt = 2\pi$ is both the distance traveled and the change in position of the object on the interval $0 \leq t \leq 4$. Specifically, the object moved 2π meters in 4 minutes.
- We know $v_{\text{AVG}}[a, b] = \frac{1}{b-a} \int_a^b v(t) dt$, so

$$v_{\text{AVG}}[0, 4] = \frac{1}{4-0} \int_0^4 \sqrt{4 - (t - 2)^2} dt = \frac{1}{4} 2\pi = \frac{\pi}{2},$$

which is measured in meters per minute, since the units of “4” are minutes and of “ 2π ” are meters.

- Constructing a figure similar to those shown in this section on the topic of average value of a function, we find the following, which demonstrates a rectangle having the same area as the semicircle.



The height of the rectangle is the average value of v , specifically $v_{\text{AVG}}[0, 4] = \frac{\pi}{2} \approx 1.57$.

- f. Knowing that average velocity is $\frac{\pi}{2}$, it follows from the fact that distance traveled equals average rate times time (provided velocity is always nonnegative), we have

$$D = \frac{\pi}{2} \cdot 4 = 2\pi.$$

We see from (c) or (f) that we are simply considering the situation from two different perspectives: if we know the distance traveled, we can find average velocity, or if we know average velocity, we can find distance traveled.

4.3.4 Summary

- 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 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

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \Delta x,$$

where $\Delta x = \frac{b-a}{n}$, $x_i = a + i\Delta x$ (for $i = 0, \dots, n$), and x_i^* satisfies $x_{i-1} \leq x_i^* \leq x_i$ (for $i = 1, \dots, n$).

- 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 addition, the value of the definite integral is related to what we call the average value of the function on $[a, b]$: $f_{\text{AVG}}[a, b] = \frac{1}{b-a} \cdot \int_a^b f(x) dx$. 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.

4.3.5 Exercises

1. (Evaluating definite integrals from graphical information) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Estimating definite integrals from a graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Finding the average value of a linear function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Finding the average value of a function given graphically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Estimating a definite integral and average value from a graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Using rules to combine known integral values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. The velocity of an object moving along an axis is given by the piecewise linear function v that is pictured in Figure 4.3.16. 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 $t = 4$.

- Write an expression involving definite integrals whose value is the total change in position of the object on the interval $[0, 4]$.
- Use the provided graph of v to determine the value of the total change in position on $[0, 4]$.
- 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 $[0, 4]$?
- What is the object's exact average velocity on $[0, 4]$?
- Find an algebraic formula for the object's position function on $[0, 1.5]$ that satisfies $s(0) = 0$.

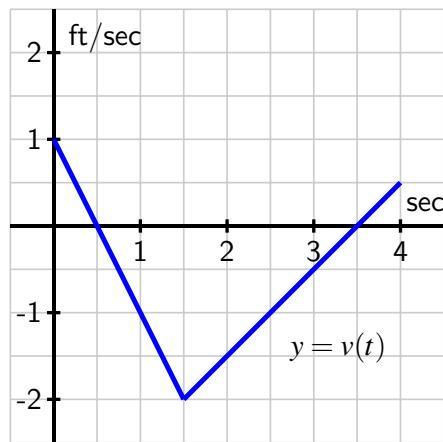


Figure 4.3.16: The velocity function of a moving object.

Answer.

- The total change in position is $P = \int_0^4 v(t)dt$.
- $P = -2.625$ feet.
- $D = 3.375$ feet.
- $AV = -0.65625$ feet per second.
- $s(t) = -t^2 + t$.

Solution.

- Since the area between the velocity function and the t -axis tells us distance traveled (provided we are careful about velocity being negative), we can use definite integrals to express the total change in position.

The change in position on $[0, 0.5]$ is $\int_0^{0.5} v(t)dt$, and the value of this integral is positive since velocity is positive. The change in position on $[0.5, 3.5]$ is $\int_{0.5}^{3.5} v(t)dt$, and the value of this integral is negative since the velocity is negative. Finally, the change in position on $[3.5, 4]$ is $\int_{3.5}^4 v(t)dt$, which is positive. Thus, the total change in position, P , is

$$P = \int_0^{0.5} v(t)dt + \int_{0.5}^{3.5} v(t)dt + \int_{3.5}^4 v(t)dt.$$

By the additive property of definite integrals, the total change in position is

$$P = \int_0^4 v(t)dt.$$

- b. We can evaluate the three integrals identified in (a) by finding the areas of relevant triangles since the velocity function is piecewise linear. In particular,

$$\int_0^{0.5} v(t)dt = \frac{1}{2}(0.5)(1) = 0.25,$$

$$\int_{0.5}^{3.5} v(t)dt = -\frac{1}{2}(3)(2) = -3,$$

and

$$\int_{3.5}^4 v(t)dt = \frac{1}{2}(0.5)(0.5) = 0.125.$$

Thus, the total change in position is

$$\begin{aligned} P &= \int_0^4 v(t)dt \\ &= \int_0^{0.5} v(t)dt + \int_{0.5}^{3.5} v(t)dt + \int_{3.5}^4 v(t)dt \\ &= 0.25 - 3 + 0.125 \\ &= -2.625 \end{aligned}$$

feet.

- c. For the total distance traveled, D , we take the opposite of the change in position on any interval where the velocity is negative. Hence,

$$\begin{aligned} D &= \int_0^{0.5} v(t)dt - \int_{0.5}^{3.5} v(t)dt + \int_{3.5}^4 v(t)dt \\ &= 0.25 + 3 + 0.125 \\ &= 3.375 \end{aligned}$$

feet.

- d. Recall that the average value of a function f on an interval $[a, b]$ is

$$f_{\text{AVG}[a,b]} = \frac{1}{b-a} \int_a^b f(x)dx.$$

So the average velocity, AV , of the object on the interval $[0, 4]$ is

$$AV = \frac{1}{4} \int_0^4 v(t)dt = (0.25)(-2.625) = -0.65625$$

feet per second.

- e. Assuming that v is linear on the interval $[0, 1.5]$, the slope of v is $-\frac{3}{1.5} = -2$. Given that the v -intercept is $(0, 1)$ we conclude that on the interval $[0, 1.5]$ we have $v(t) = -2t + 1$. Now position, s , is an antiderivative of the velocity, v , so the possible formulas for s on the interval $[0, 1.5]$ are $s(t) = -t^2 + t + C$, where C is some constant. If we want $s(0) = 0$, that forces $C = 0$ and so the object's position function on $[0, 1.5]$ that satisfies $s(0) = 0$ is $s(t) = -t^2 + t$.

8. Suppose that the velocity of a moving object is given by $v(t) = t(t - 1)(t - 3)$, measured in feet per second, and that this function is valid for $0 \leq t \leq 4$.

- Write an expression involving definite integrals whose value is the total change in position of the object on the interval $[0, 4]$.
- Use appropriate technology (such as <http://gvsu.edu/s/a94>) to compute Riemann sums to estimate the object's total change in position on $[0, 4]$. Work to ensure that your estimate is accurate to two decimal places, and explain how you know this to be the case.
- Write an expression involving definite integrals whose value is the total distance traveled by the object on $[0, 4]$.
- Use appropriate technology to compute Riemann sums to estimate the object's total distance travelled on $[0, 4]$. Work to ensure that your estimate is accurate to two decimal places, and explain how you know this to be the case.
- What is the object's average velocity on $[0, 4]$, accurate to two decimal places?

Answer.

- The total change in position, P , is $P = \int_0^1 v(t) dt + \int_1^3 v(t) dt + \int_3^4 v(t) dt = \int_0^4 v(t) dt$.
- $P = \int_0^4 v(t) dt \approx 2.665$.
- The total distance traveled, D , is $D = \int_0^1 v(t) dt - \int_1^3 v(t) dt + \int_3^4 v(t) dt$.
- $D \approx 8.00016$.
- e.

$$v_{\text{AVG}[0,4]} \approx 0.66625$$

feet per second.

Solution.

- We observe that $v(t) = 0$ at $t = 0$, $t = 1$, and $t = 3$, and $v(t)$ is positive on the intervals $0 < t < 1$ and $3 < t < 4$, while $v(t)$ is negative on $1 < t < 3$. The total change in position, P , of the object is thus given by

$$P = \int_0^1 v(t) dt + \int_1^3 v(t) dt + \int_3^4 v(t) dt = \int_0^4 v(t) dt.$$

- Using computational technology, we use the midpoint rule to estimate the total change of position, $\int_0^4 v(t) dt$. We find that $M_{25} = 2.6496$, $M_{50} = 2.6624$, $M_{75} = 2.66477$, and $M_{100} = 2.6656$. Since it appears that neither of the first two decimal places are changing as we increase the number of subintervals, it looks like $P = \int_0^4 v(t) dt \approx 2.665$ is accurate to two decimal places.
- We noted in (a) where $v(t)$ is positive and negative. In particular, on $1 < t < 3$, $v(t) < 0$, and thus $\int_1^3 v(t) dt$ will give us a negative change in position; the distance traveled on this interval is therefore $-\int_1^3 v(t) dt$, and hence the total distance traveled, D , on $[0, 4]$ is

$$D = \int_0^1 v(t) dt - \int_1^3 v(t) dt + \int_3^4 v(t) dt.$$

⁴Marc Renault, Shippensburg University.

- d. To use Riemann sums to estimate D , we have to estimate each of the three integrals individually. We find that

- $\int_0^1 v(t) dt \approx M_{100} = 0.41669$,
- $\int_1^3 v(t) dt \approx M_{100} = -2.6668$, and
- $\int_3^4 v(t) dt \approx M_{100} = 4.91667$

From experimenting with different n -values in M_n , it is apparent that each of these individual estimates is accurate to at least 3 decimal places, and thus we expect their sum to be accurate to at least 2. Hence,

$$D = \int_0^1 v(t) dt - \int_1^3 v(t) dt + \int_3^4 v(t) dt \approx 0.41669 - (-2.6668) + 4.91667 = 8.00016.$$

Therefore $D \approx 8.00016$ is an estimate of distance traveled accurate to at least 2 decimal places.

- e. We know that the average velocity on $[0, 4]$ is given by the integral expression

$$v_{\text{AVG}[0,4]} = \frac{1}{4-0} \int_0^4 v(t) dt.$$

We estimated the definite integral in (b), so using that result,

$$v_{\text{AVG}[0,4]} \approx \frac{1}{4-0} 2.665 = 0.66625$$

feet per second is the object's average velocity, accurate to at least two decimal places. It makes sense that the average velocity is positive since the total change in position is positive.

9. Consider the graphs of two functions f and g that are provided in Figure 4.3.17. Each piece of f and g is either part of a straight line or part of a circle.

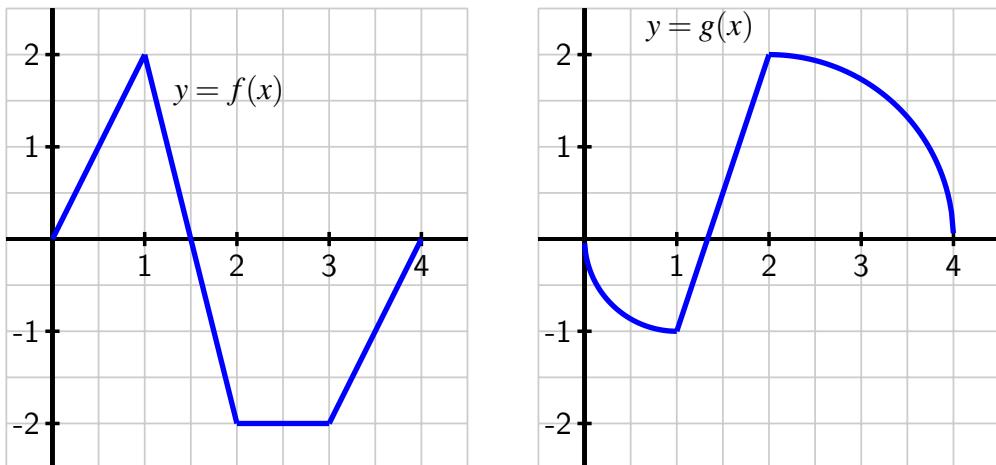


Figure 4.3.17: Two functions f and g .

- Determine the exact value of $\int_0^1 [f(x) + g(x)] dx$.
- Determine the exact value of $\int_1^4 [2f(x) - 3g(x)] dx$.
- Find the exact average value of $h(x) = g(x) - f(x)$ on $[0, 4]$.
- For what constant c does the following equation hold?

$$\int_0^4 c dx = \int_0^4 [f(x) + g(x)] dx$$

Answer.

- $\int_0^1 [f(x) + g(x)] dx = 1 - \frac{\pi}{4}$.
- $\int_1^4 [2f(x) - 3g(x)] dx = -\frac{15}{2} - 3\pi$.
- $h_{\text{AVG}[0,4]} = \frac{5}{8} + \frac{3\pi}{16}$.
- $c = -\frac{3}{8} + \frac{3\pi}{16}$.

Solution.

- The sum property of the definite integral tells us that

$$\int_0^1 [f(x) + g(x)] dx = \int_0^1 f(x) dx + \int_0^1 g(x) dx$$

Now $\int_0^1 f(x) dx$ is the area of a triangle of height 2 and base 1, so $\int_0^1 f(x) dx = 1$. Also, $\int_0^1 g(x) dx$ is the opposite of the area of a quarter circle of radius 1, so $\int_0^1 g(x) dx = -\frac{\pi}{4}$. Therefore,

$$\int_0^1 [f(x) + g(x)] dx = \int_0^1 f(x) dx + \int_0^1 g(x) dx = 1 - \frac{\pi}{4}.$$

- The value of $\int_1^4 f(x) dx$ is the difference of the areas of the regions f bounds above the x -axis minus the areas of the regions f bounds below the x -axis on the interval $[1, 4]$. The area of the triangle f bounds above the x -axis on the interval $[1, 1.5]$ is $(0.5)(2)(0.5) = 0.5$, the area of the region f bounds below the x -axis on $[1.5, 2]$ is also 0.5, the area of the region f bounds below the x -axis on $[2, 3]$ is the area (2) of a rectangle with height 2 and base 1, and the area of the region f bounds below the x -axis on $[3, 4]$ is the area (1) of a triangle with height 2 and base 1. So

$$\int_1^4 f(x) dx = 0.5 - 0.5 - 2 - 1 = -3.$$

Notice that the graph of g is linear on the interval $[1, 2]$, passing through the points $(1, -1)$ and $(2, 2)$. This line has slope 3 and equation $y = -1 + 3(x - 1)$. The x -intercept of this line is $\frac{4}{3}$. So the triangle the graph of g bounds on the interval $[1, \frac{4}{3}]$ has area $(\frac{1}{2})(\frac{1}{3})(1) = \frac{1}{6}$, and the area of the triangle between the graph of g and the x -axis on the interval $[\frac{4}{3}, 2]$ is $(\frac{1}{2})(\frac{2}{3})(2) = \frac{2}{3}$. Finally, the graph of g bounds a quarter circle of radius 2 on the interval $[2, 4]$, so

$$\int_1^4 g(x) dx = -\frac{1}{6} + \frac{2}{3} + \pi = \frac{1}{2} + \pi.$$

Therefore,

$$\int_1^4 [2f(x) - 3g(x)] dx = 2 \int_1^4 f(x) dx - 3 \int_1^4 g(x) dx = (2)(-3) - (3) \left(\frac{1}{2} + \pi \right) = -\frac{15}{2} - 3\pi.$$

- c. Recall that the average value of a function f on an interval $[a, b]$ is

$$f_{\text{AVG}[a,b]} = \frac{1}{b-a} \int_a^b f(x) dx.$$

Combining the results from (a) and (b) gives us

$$\int_0^4 f(x) dx = \int_0^1 f(x) dx + \int_1^4 f(x) dx = 1 - 3 = -2$$

and

$$\int_0^4 g(x) dx = \int_0^1 g(x) dx + \int_1^4 g(x) dx = -\frac{\pi}{4} + \frac{1}{2} + \pi = \frac{1}{2} + \frac{3\pi}{4}.$$

Thus, the average value of h on the interval $[0, 4]$ is

$$\begin{aligned} h_{\text{AVG}[0,4]} &= \frac{1}{4} \int_0^4 [g(x) - f(x)] dx \\ &= \frac{1}{4} \left(\int_0^4 g(x) dx - \int_0^4 f(x) dx \right) \\ &= \frac{1}{4} \left(\frac{1}{2} + \frac{3\pi}{4} - (-2) \right) \\ &= \frac{5}{8} + \frac{3\pi}{16}. \end{aligned}$$

- d. We know that $\int_0^4 c dx$ is the signed area of a rectangle with “height” c (note that c could be negative), and base 4, so

$$\int_0^4 c dx = 4c.$$

Also, from part (c) we have

$$\int_0^4 [f(x) + g(x)] dx = \int_0^4 f(x) dx + \int_0^4 g(x) dx = -2 + \frac{1}{2} + \frac{3\pi}{4} = -\frac{3}{2} + \frac{3\pi}{4}.$$

Therefore,

$$\int_0^4 c dx = \int_0^4 [f(x) + g(x)] dx$$

from which it follows that $4c = -\frac{3}{2} + \frac{3\pi}{4}$ so $c = -\frac{3}{8} + \frac{3\pi}{16}$.

10. Let $f(x) = 3 - x^2$ and $g(x) = 2x^2$.

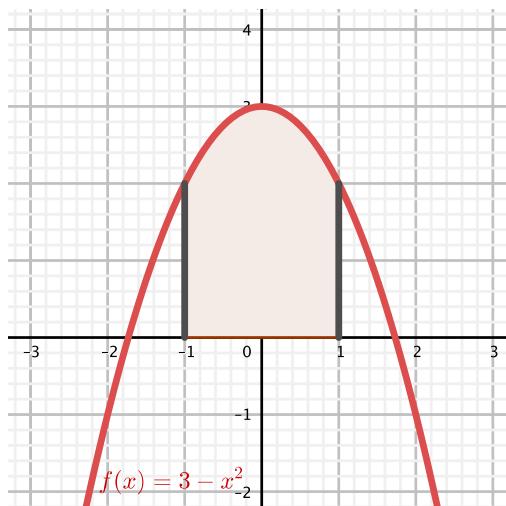
- a. On the interval $[-1, 1]$, sketch a labeled graph of $y = f(x)$ and write a definite integral whose value is the exact area bounded by $y = f(x)$ on $[-1, 1]$.

- b. On the interval $[-1, 1]$, sketch a labeled graph of $y = g(x)$ and write a definite integral whose value is the exact area bounded by $y = g(x)$ on $[-1, 1]$.
- c. Write an expression involving a difference of definite integrals whose value is the exact area that lies between $y = f(x)$ and $y = g(x)$ on $[-1, 1]$.
- d. Explain why your expression in (c) has the same value as the single integral $\int_{-1}^1 [f(x) - g(x)] dx$.
- e. Explain why, in general, if $p(x) \geq q(x)$ for all x in $[a, b]$, the exact area between $y = p(x)$ and $y = q(x)$ is given by

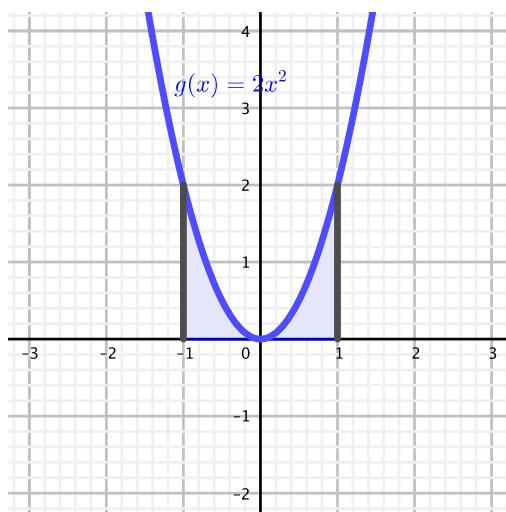
$$\int_a^b [p(x) - q(x)] dx.$$

Answer.

a. $A_1 = \int_{-1}^1 (3 - x^2) dx$.



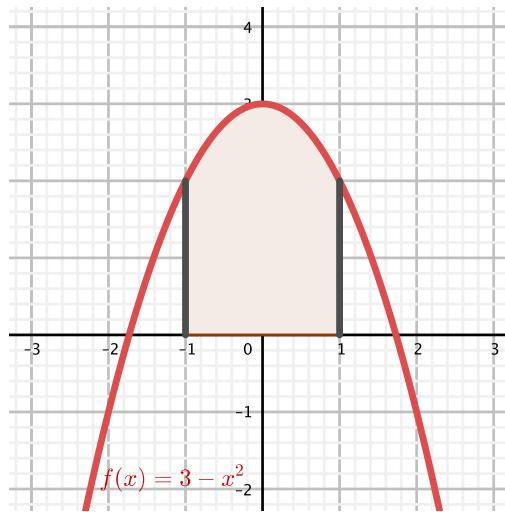
b. $A_2 = \int_{-1}^1 2x^2 dx$.



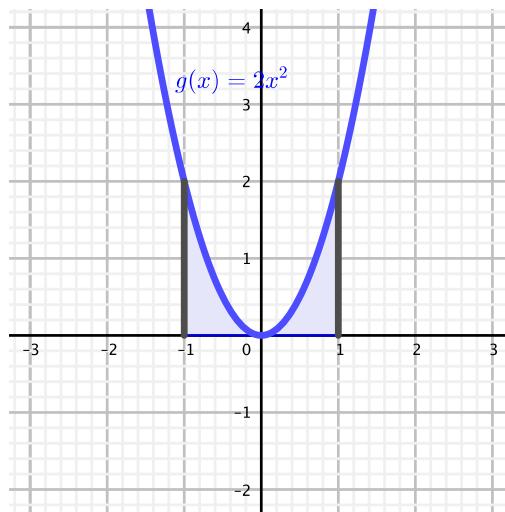
- c. The exact area between the two curves is $\int_{-1}^1 (3 - x^2) dx - \int_{-1}^1 2x^2 dx$.
- d. Use the sum rule for definite integrals over the same interval.
- e. Think about subtracting the area under q from the area under p .

Solution.

- a. Since $f(x)$ is positive on $[-1, 1]$, the area bounded by the function and x -axis on the given interval is $A_1 = \int_{-1}^1 (3 - x^2) dx$. A sketch of the function and region is shown below.



- b. Likewise, because $g(x)$ is positive on $[-1, 1]$, the area bounded by the function and x -axis on the given interval is $A_2 = \int_{-1}^1 2x^2 dx$. A sketch of the function and region is shown below.



- c. From the above two figures, we see that the exact area between the two curves is the

area under f (A_1) minus the area under g (A_2). That is,

$$A_1 - A_2 = \int_{-1}^1 (3 - x^2) dx - \int_{-1}^1 2x^2 dx.$$

d. By the sum rule for definite integrals over the same interval, we know that

$$\int_{-1}^1 (3 - x^2) dx - \int_{-1}^1 2x^2 dx = \int_{-1}^1 [(3 - x^2) - 2x^2] dx$$

and thus the latter single integral captures the exact area between the two curves on $[-1, 1]$.

e. In general, if $p(x) \geq q(x)$ on $[a, b]$, the exact area, A , between p and q on the interval will be given by

$$A = \int_a^b [p(x) - q(x)] dx$$

because we can find the area as the difference of the area under p ($\int_a^b p(x) dx$) and the area under q ($\int_a^b q(x) dx$), and then subtract the two definite integrals as shown in (d).

4.4 The Fundamental Theorem of Calculus

Motivating 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, 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?
-

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 the distance it traveled on $[a, b]$? If the velocity function is nonnegative on $[a, b]$, the area bounded by $y = v(t)$ and the t -axis on $[a, b]$ is equal to the distance traveled. This area is also the value of the definite integral $\int_a^b v(t) dt$. If the velocity is sometimes negative, the total area bounded by the velocity function still tells us distance traveled, while the net signed area tells us the object's change in position.

For instance, for the velocity function in Figure 4.4.1, the total distance D traveled by the moving object on $[a, b]$ is

$$D = A_1 + A_2 + A_3,$$

and the total change in the object's position is

$$s(b) - s(a) = A_1 - A_2 + A_3.$$

The areas A_1 , A_2 , and A_3 are each given by definite integrals, which may be computed by limits of Riemann sums (and in special circumstances by geometric formulas).

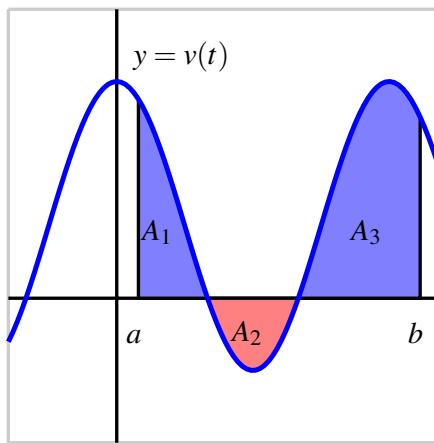


Figure 4.4.1: A velocity function that is sometimes negative.

We turn our attention to an alternate approach.

Example 4.4.2. 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 $v(t) = -32t + 16$, where v is measured in feet per second and t is measured in seconds.

- Let $s(t)$ represent the height of the water balloon above ground at time t , and note that s is an antiderivative of v . That is, v is the derivative of s : $s'(t) = v(t)$. 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 $s(0) = 32$.
- When does the water balloon reach its maximum height? When does it land?
- Compute $s(\frac{1}{2}) - s(0)$, $s(2) - s(\frac{1}{2})$, and $s(2) - s(0)$. What do these represent?
- What is the total vertical distance traveled by the water balloon from the time it is tossed until the time it lands?
- 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.

4.4.1 The Fundamental Theorem of Calculus

Suppose we know the position function $s(t)$ and the velocity function $v(t)$ of an object moving in a straight line, and for the moment let us assume that $v(t)$ is positive on $[a, b]$. Then, as shown in Figure 4.4.3, we know two different ways to compute the distance, D , the object

travels: one is that $D = s(b) - s(a)$, the object's change in position. The other is the area under the velocity curve, which is given by the definite integral, so $D = \int_a^b v(t) dt$.

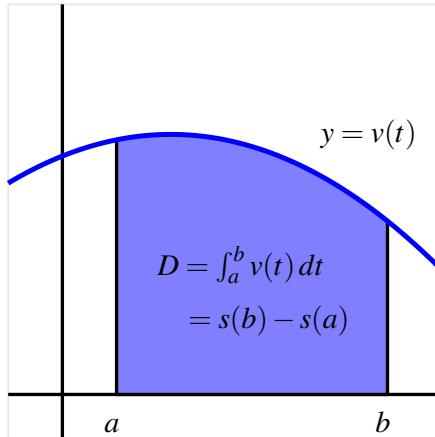


Figure 4.4.3: Finding distance traveled when we know a velocity function v .

Since both of these expressions tell us the distance traveled, it follows that they are equal, so

$$s(b) - s(a) = \int_a^b v(t) dt. \quad (4.4.1)$$

Equation (4.4.1) holds even when velocity is sometimes negative, because $s(b) - s(a)$, the object's change in position, is also measured by the net signed area on $[a, b]$ which is given by $\int_a^b v(t) dt$.

Perhaps the most powerful fact Equation (4.4.1) reveals is 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 .

Example 4.4.4. Determine the exact distance traveled on $[1, 5]$ by an object with velocity function $v(t) = 3t^2 + 40$ feet per second. The distance traveled on the interval $[1, 5]$ is given by

$$D = \int_1^5 v(t) dt = \int_1^5 (3t^2 + 40) dt = s(5) - s(1),$$

where s is an antiderivative of v . Now, the derivative of t^3 is $3t^2$ and the derivative of $40t$ is 40, so it follows that $s(t) = t^3 + 40t$ is an antiderivative of v . Therefore,

$$\begin{aligned} D &= \int_1^5 3t^2 + 40 dt = s(5) - s(1) \\ &= (5^3 + 40 \cdot 5) - (1^3 + 40 \cdot 1) = 284 \text{ feet.} \end{aligned}$$

Note the key lesson of Example 4.4.4: to find the distance traveled, we need to compute the area under a curve, which is given by the definite integral. But to evaluate the integral, we can find an antiderivative, s , of the velocity function, and then compute the total change in s on the interval. In particular, we can evaluate the integral without computing the limit of a Riemann sum.

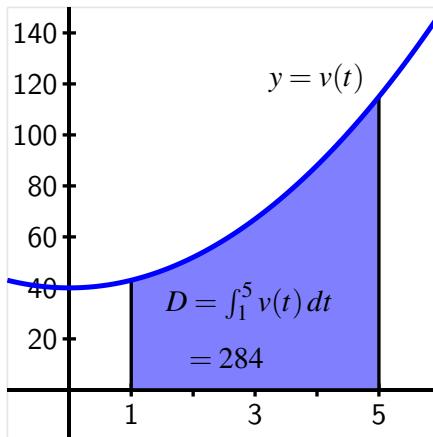


Figure 4.4.5: The exact area of the region enclosed by $v(t) = 3t^2 + 40$ on $[1, 5]$.

It will be convenient to have a shorthand symbol for a function's antiderivative. For a continuous function f , we will often denote an antiderivative of f by F , so 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 (4.4.1), we can write

$$V(b) - V(a) = \int_a^b v(t) dt. \quad (4.4.2)$$

Now, 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. But Equations (4.4.1) and (4.4.2) hold for any continuous velocity function, even when v is sometimes negative. So Equation (4.4.2) offers a shortcut route to evaluating any definite integral, provided that we can find an antiderivative of the integrand. The Fundamental Theorem of Calculus (FTC) summarizes these observations.

Fundamental Theorem of Calculus

If f is a continuous function on $[a, b]$, and F is any antiderivative of f , then $\int_a^b f(x) dx = F(b) - F(a)$.

A common alternate notation for $F(b) - F(a)$ is

$$F(b) - F(a) = F(x)|_a^b,$$

where we read the righthand side as “the function F evaluated from a to b .” In this notation, the FTC says that

$$\int_a^b f(x) dx = F(x)|_a^b.$$

The FTC opens the door to evaluating a wide range of integrals if we can find an antiderivative F for the integrand f . For instance since $\frac{d}{dx}[\frac{1}{3}x^3] = x^2$, the FTC tells us that

$$\begin{aligned}\int_0^1 x^2 dx &= \frac{1}{3} x^3 \Big|_0^1 \\ &= \frac{1}{3} (1)^3 - \frac{1}{3} (0)^3 \\ &= \frac{1}{3}.\end{aligned}$$

But finding an antiderivative can be far from simple; it is often difficult or even impossible. While we can differentiate just about any function, even some relatively simple functions 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 the problem of finding antiderivatives.

Example 4.4.6. Use the Fundamental Theorem of Calculus to evaluate each of the following integrals exactly. For each, sketch a graph of the integrand on the relevant interval and write one sentence that explains the meaning of the value of the integral in terms of the (net signed) area bounded by the curve.

- | | |
|--|--------------------------------------|
| a. $\int_{-1}^4 (2 - 2x) dx$ | d. $\int_{-1}^1 x^5 dx$ |
| b. $\int_0^{\frac{\pi}{2}} \sin(x) dx$ | |
| c. $\int_0^1 e^x dx$ | e. $\int_0^2 (3x^3 - 2x^2 - e^x) dx$ |

Hint.

- a. Find a function whose derivative is $2 - 2x$.
- b. Which familiar function has derivative $\sin(x)$?
- c. What is special about e^x when it comes to differentiation?
- d. Consider the derivative of x^6 .
- e. Find an antiderivative for each of the three individual terms in the integrand.

Answer.

- a. $\int_{-1}^4 (2 - 2x) dx = -5.$
- b. $\int_0^{\frac{\pi}{2}} \sin(x) dx = 1.$

- c. $\int_0^1 e^x dx = e - 1$.
- d. $\int_{-1}^1 x^5 dx = 0$.
- e. $\int_0^2 (3x^3 - 2x^2 - e^x) dx = \frac{17}{3} - e^2$.

Solution.

- a. Because $\frac{d}{dx}[2x - x^2] = 2 - 2x$, by the Fundamental Theorem of Calculus,

$$\int_{-1}^4 (2 - 2x) dx = 2x - x^2 \Big|_{-1}^4,$$

and therefore

$$\int_{-1}^4 (2 - 2x) dx = (2 \cdot 4 - 4^2) - (2(-1) - (-1)^2) = -8 + 3 = -5.$$

- b. Since $\frac{d}{dx}[\cos(x)] = -\sin(x)$, an antiderivative of $f(x) = \sin(x)$ is $F(x) = -\cos(x)$. Therefore, by the FTC,

$$\int_0^{\frac{\pi}{2}} \sin(x) dx = -\cos(x) \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}},$$

so

$$\int_0^{\frac{\pi}{2}} \sin(x) dx = -\cos\left(\frac{\pi}{2}\right) - (-\cos(0)) = -0 + 1 = 1.$$

- c. Since e^x is its own derivative, it is also its own antiderivative. Hence,

$$\int_0^1 e^x dx = e^x \Big|_0^1 = e^1 - e^0 = e - 1.$$

- d. Note that since $\frac{d}{dx}[x^6] = 6x^5$, it follows that $\frac{d}{dx}\left[\frac{1}{6}x^6\right] = x^5$, and thus

$$\int_{-1}^1 x^5 dx = \frac{1}{6}x^6 \Big|_{-1}^1 = \frac{1}{6}(1)^6 - \frac{1}{6}(-1)^6 = 0.$$

- e. Using the sum and constant multiple rules for differentiation, we can see that similar results hold for antiderivation, and thus that $F(x) = \frac{3}{4}x^4 - \frac{2}{3}x^3 - e^x$ is an antiderivative of $f(x) = 3x^3 - 2x^2 - e^x$. Now, by the FTC,

$$\begin{aligned} \int_0^2 (3x^3 - 2x^2 - e^x) dx &= \frac{3}{4}x^4 - \frac{2}{3}x^3 - e^x \Big|_0^2 \\ &= \frac{3}{4}(2)^4 - \frac{2}{3}(2)^3 - e^2 - \left(\frac{3}{4}(0)^4 - \frac{2}{3}(0)^3 - e^0\right) \\ &= 12 - \frac{16}{3} - e^2 - (0 - 0 - 1) \\ &= \frac{17}{3} - e^2. \end{aligned}$$

4.4.2 The total change theorem

Let us review three interpretations of the definite integral.

- For a moving object with instantaneous velocity $v(t)$, the object's change in position on the time interval $[a, b]$ is given by $\int_a^b v(t) dt$, and whenever $v(t) \geq 0$ on $[a, b]$, $\int_a^b v(t) dt$ tells us the total distance traveled by the object on $[a, b]$.
- For any continuous function f , its definite integral $\int_a^b f(x) dx$ represents the net signed area bounded by $y = f(x)$ and the x -axis on $[a, b]$, where regions that lie below the x -axis have a minus sign associated with their area.
- The value of a definite integral is linked to the average value of a function: for a continuous function f on $[a, b]$, its average value $f_{\text{AVG}}[a, b]$ is given by

$$f_{\text{AVG}}[a, b] = \frac{1}{b - a} \int_a^b f(x) dx.$$

The Fundamental Theorem of Calculus now enables us to evaluate exactly (without taking a limit of Riemann sums) any definite integral for which we are able to find an antiderivative of the integrand.

A slight change in perspective allows us to gain even more insight into the meaning of the definite integral. Recall Equation (4.4.2), where we wrote the Fundamental Theorem of Calculus for a velocity function v with antiderivative V as

$$V(b) - V(a) = \int_a^b v(t) dt.$$

If we instead replace V with s (which represents position) and replace v with s' (since velocity is the derivative of position), Equation (4.4.2) then reads as

$$s(b) - s(a) = \int_a^b s'(t) dt. \quad (4.4.3)$$

In words, this version of the FTC tells us that the total change in an object's position function on a particular interval is given by the definite integral of the position function's derivative over that interval.

Of course, this result is not limited to only the setting of position and velocity. Writing the result in terms of a more general function f , we have the Total Change Theorem.

Total Change Theorem

If f is a continuously differentiable function on $[a, b]$ with derivative f' , then $f(b) - f(a) = \int_a^b f'(x) dx$. That is, the definite integral of the rate of change of a function on $[a, b]$ is the total change of the function itself on $[a, b]$.

The Total Change Theorem tells us more about the relationship between the graph of a function and that of its derivative. Recall that heights on the graph of the derivative function are

equal to slopes on the graph of the function itself. If instead we know f' and are seeking information about f , we can say the following:

differences in heights on f correspond to net signed areas bounded by f' .

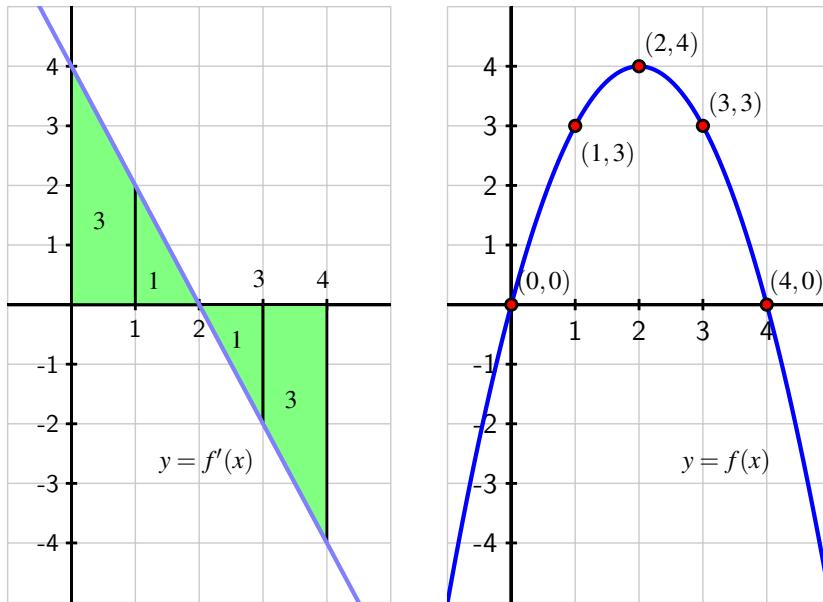


Figure 4.4.7: The graphs of $f'(x) = 4 - 2x$ (at left) and an antiderivative $f(x) = 4x - x^2$ at right. Differences in heights on f correspond to net signed areas bounded by f' .

To see why this is so, consider the difference $f(1) - f(0)$. This value is 3, because $f(1) = 3$ and $f(0) = 0$, but also because the net signed area bounded by $y = f'(x)$ on $[0, 1]$ is 3. That is,

$$f(1) - f(0) = \int_0^1 f'(x) dx.$$

In addition to this observation about area, the Total Change Theorem enables us to answer questions about a function whose rate of change we know.

Example 4.4.8. Suppose that pollutants are leaking out of an underground storage tank at a rate of $r(t)$ gallons/day, where t is measured in days. It is conjectured that $r(t)$ is given by the formula $r(t) = 0.0069t^3 - 0.125t^2 + 11.079$ over a certain 12-day period. The graph of $y = r(t)$ is given in Figure 4.4.9. What is the meaning of $\int_4^{10} r(t) dt$ and what is its value? What is the average rate at which pollutants are leaving the tank on the time interval $4 \leq t \leq 10$?

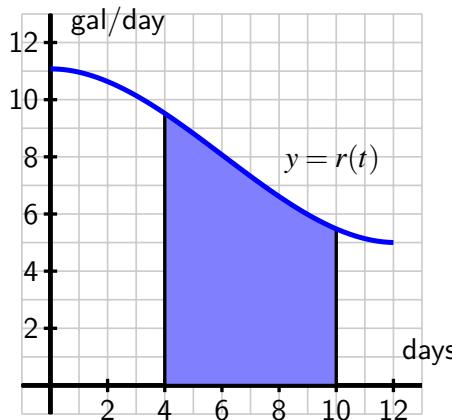


Figure 4.4.9: The rate $r(t)$ of pollution leaking from a tank, measured in gallons per day.

Solution. Since $r(t) \geq 0$, the value of $\int_4^{10} r(t) dt$ is the area under the curve on the interval $[4, 10]$. A Riemann sum for this area will have rectangles with heights measured in gallons per day and widths measured in days, so the area of each rectangle will have units of

$$\frac{\text{gallons}}{\text{day}} \cdot \text{days} = \text{gallons}.$$

Thus, the definite integral tells us the total number of gallons of pollutant that leak from the tank from day 4 to day 10. The Total Change Theorem tells us the same thing: if we let $R(t)$ denote the total number of gallons of pollutant that have leaked from the tank up to day t , then $R'(t) = r(t)$, and

$$\int_4^{10} r(t) dt = R(10) - R(4),$$

the number of gallons that have leaked from day 4 to day 10.

To compute the exact value of the integral, we use the Fundamental Theorem of Calculus. Antidifferentiating $r(t) = 0.0069t^3 - 0.125t^2 + 11.079$, we find that

$$\begin{aligned} \int_4^{10} 0.0069t^3 - 0.125t^2 + 11.079 dt &= 0.0069 \cdot \frac{1}{4} t^4 - 0.125 \cdot \frac{1}{3} t^3 + 11.079t \Big|_4^{10} \\ &\approx 44.282. \end{aligned}$$

Thus, approximately 44.282 gallons of pollutant leaked over the six day time period.

To find the average rate at which pollutant leaked from the tank over $4 \leq t \leq 10$, we compute the average value of r on $[4, 10]$. Thus,

$$r_{\text{AVG}[4,10]} = \frac{1}{10 - 4} \int_4^{10} r(t) dt \approx \frac{44.282}{6} = 7.380$$

gallons per day.

Example 4.4.10. During a 40-minute workout, a person riding an exercise machine burns calories at a rate of c calories per minute, where the function $y = c(t)$ is given in Figure 4.4.11. On the interval $0 \leq t \leq 10$, the formula for c is $c(t) = -0.05t^2 + t + 10$, while on $30 \leq t \leq 40$, its formula is $c(t) = -0.05t^2 + 3t - 30$.

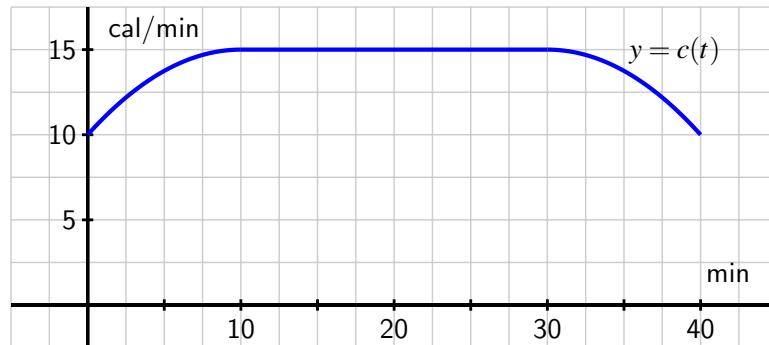


Figure 4.4.11: The rate $c(t)$ at which a person exercising burns calories, measured in calories per minute.

- What is the exact total number of calories the person burns during the first 10 minutes of her workout?
- Let $C(t)$ be an antiderivative of $c(t)$. What is the meaning of $C(40) - C(0)$ in the context of the person exercising? Include units in your answer.
- Determine the exact average rate at which the person burned calories during the 40-minute workout.
- At what time(s), if any, is the instantaneous rate at which the person is burning calories equal to the average rate at which she burns calories, on the time interval $0 \leq t \leq 40$?

Hint.

- What are the units of the area of a rectangle found in a Riemann sum for the function $y = c(t)$?
- Use the FTC.
- Recall the formula for $c_{\text{AVG}[0,40]}$.
- Think carefully about which function tells you the instantaneous rate at which calories are burned.

Answer.

- a. The person burned approximately 133.33 calories in the first 10 minutes of the workout.
- b. $C(40) - C(0) = \int_0^{40} C'(t) dt = \int_0^{40} c(t) dt$ is the total calories burned on $[0, 40]$.
- c. The exact average rate at which the person burned calories on $0 \leq t \leq 40$ is

$$c_{\text{AVG}[0,40]} = \frac{1}{40-0} \int_0^{40} c(t) dt = \frac{1}{40} \cdot \frac{1700}{3} = \frac{1700}{120} \approx 14.17 \text{ cal/min.}$$

- d. One time at which the instantaneous rate at which calories are burned equals the average rate on $[0, 40]$ is $t = \frac{5}{3}(6 - \sqrt{6}) \approx 5.918$.

Solution.

- a. Since the units on a rectangle in a Riemann sum are cal/min for the height and min for the width, the units of the area of such a rectangle are calories, and hence the units of area under the curve $y = c(t)$ are given in total calories. Hence, the total calories burned during the first 10 minutes of the workout is given by the definite integral $\int_0^{10} c(t) dt$. We use the FTC and evaluate the integral, finding that

$$\begin{aligned} \int_0^{10} (-0.05t^2 + t + 10) dt &= \left(-\frac{0.05}{3}t^3 + \frac{1}{2}t^2 + 10t \right) \Big|_0^{10} \\ &= \left(-\frac{0.05}{3}(10)^3 + \frac{1}{2}(10)^2 + 10(10) \right) - (-0 + 0 + 0) \\ &= \frac{400}{3} \text{ calories.} \end{aligned}$$

Thus, the person burned approximately 133.33 calories in the first 10 minutes of the workout.

- b. We observe first that by the Total Change Theorem, $C(40) - C(0) = \int_0^{40} C'(t) dt = \int_0^{40} c(t) dt$, and therefore, as discussed in (a), the meaning of this value is the total calories burned on $[0, 40]$.
- c. The exact average rate at which the person burned calories on $0 \leq t \leq 40$ is given by

$$c_{\text{AVG}[0,40]} = \frac{1}{40-0} \int_0^{40} c(t) dt.$$

To calculate $\int_0^{40} c(t) dt$, we recognize that $c(t)$ is defined in piecewise fashion, and use the additive property of the definite integral, which tells us that

$$\int_0^{40} c(t) dt = \int_0^{10} c(t) dt + \int_{10}^{30} c(t) dt + \int_{30}^{40} c(t) dt.$$

We know from our work in (a) that $\int_0^{10} c(t) dt = \frac{400}{3}$. Since $c(t) = 15$ is constant on $10 \leq t \leq 20$, it follows that $\int_{10}^{30} c(t) dt = 15 \cdot 30 = 300$. And finally, it is straightforward

to show using $c(t) = -0.05t^2 + 3t - 30$ (or symmetry) on $30 \leq t \leq 40$ that $\int_{30}^{40} c(t) dt = \frac{400}{3}$. Hence,

$$\begin{aligned}\int_0^{40} c(t) dt &= \int_0^{10} c(t) dt + \int_{10}^{30} c(t) dt + \int_{30}^{40} c(t) dt \\ &= \frac{400}{3} + 300 + \frac{400}{3} \\ &= \frac{1700}{3} \approx 566.67 \text{ calories.}\end{aligned}$$

Now, it follows that the exact average rate at which calories were burned on $[0, 40]$ is

$$c_{\text{AVG}[0,40]} = \frac{1}{40-0} \int_0^{40} c(t) dt = \frac{1}{40} \cdot \frac{1700}{3} = \frac{1700}{120} \approx 14.17 \text{ cal/min.}$$

- d. It makes sense intuitively that there must be at least one time at which the instantaneous rate at which calories are burned equals the average rate at which calories are burned, as it would be impossible for a continuous instantaneous rate of change to always be above its average value. Since we know from (c) that $c_{\text{AVG}[0,40]} = \frac{85}{6}$, and $c(t)$ tells us the instantaneous rate at which calories are burned, it follows that we want to solve the equation

$$c(t) = \frac{85}{6}.$$

From the graph, it appears that there are two such values of t for which this equation is true, one in the first ten minutes, and one in the last ten. For instance, solving

$$-0.05t^2 + t + 10 = \frac{85}{6},$$

it follows that $t = \frac{5}{3}(6 \pm \sqrt{6}) \approx 14.082, 5.918$, only the second of which lies in $0 \leq t \leq 10$. So one time at which the instantaneous rate at which calories are burned equals the average rate on $[0, 40]$ is $t = \frac{5}{3}(6 - \sqrt{6}) \approx 5.918$. Similar reasoning leads to the second time that lies in $[30, 40]$.

4.4.3 Summary

- 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 $[a, b]$.

4.4.4 Exercises

1. (Using Graphs to Evaluate) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Using Graphs to Evaluate) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Estimating using the FTC) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Finding Values using the FTC) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Average Value) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Average Value (Estimating from a Graph)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Average Value (Estimating from a Table)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. 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	6000	7000	8000	9000	10,000
c (ft/min)	925	875	830	780	730	685	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.

- 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.
- 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.
- 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.

Answer.

a.

h (feet)	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10,000
c (ft/min)	925	875	830	780	730	685	635	585	535	490	440
m (min/ft)	$\frac{1}{925}$	$\frac{1}{875}$	$\frac{1}{830}$	$\frac{1}{780}$	$\frac{1}{730}$	$\frac{1}{685}$	$\frac{1}{635}$	$\frac{1}{585}$	$\frac{1}{535}$	$\frac{1}{490}$	$\frac{1}{440}$

- b. The antiderivative function tells the total number of minutes it takes for the plane to climb to an altitude of h feet.
 c. $M = \int_0^{10000} m(h) dh$.
 d. It takes the plane about $M_5 \approx 15.27$ minutes.

Solution.

- a. Since the given climb rate, $c(h)$, is in feet per minute, and the new function, $m(h)$, is measured in minutes per foot, it follows that $m(h)$ is the reciprocal of $c(h)$, so $m(h) = \frac{1}{c(h)}$. Hence we get the following updated table.

h (feet)	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10,000
c (ft/min)	925	875	830	780	730	685	635	585	535	490	440
m (min/ft)	$\frac{1}{925}$	$\frac{1}{875}$	$\frac{1}{830}$	$\frac{1}{780}$	$\frac{1}{730}$	$\frac{1}{685}$	$\frac{1}{635}$	$\frac{1}{585}$	$\frac{1}{535}$	$\frac{1}{490}$	$\frac{1}{440}$

- b. Given an original function, we know that the units on its derivative are “units of input per unit of output.” Since the units on $m(h)$ are “minutes per foot”, it follows that the antiderivative of $m(h)$ should have output units of “minutes” since its input units are “feet”. The antiderivative function should thus tell us the total number of minutes that it takes for the plane to climb to an altitude of h feet.
 c. The number of minutes required for the airplane to ascend to 10,000 feet of altitude is given by the definite integral

$$M = \int_0^{10000} m(h) dh$$

since the units on $m(h)$ are “minutes per foot” and the units on Δh in the corresponding Riemann sum are “feet”, so when we take the product of quantities with these units, the resulting numerical quantity is measured in minutes.

- d. For the Riemann sum M_5 , we have 5 subintervals. Given the original data, this means that we need to evaluate the function at every other known data point. In particular,

$$\begin{aligned} \int_0^{10000} m(h) dh &\approx M_5 \\ &= m(1000) \cdot 2000 + m(3000) \cdot 2000 + m(5000) \cdot 2000 \\ &\quad + m(7000) \cdot 2000 + m(9000) \cdot 2000 \\ &= \left(\frac{1}{875} + \frac{1}{780} + \frac{1}{685} + \frac{1}{585} + \frac{1}{490} \right) \cdot 2000 \\ &\approx 0.007634980475 \cdot 2000 \end{aligned}$$

$$\approx 15.27.$$

Thus, it takes the plane a bit more than 15 minutes to climb to an altitude of 10,000 feet.

9. In Chapter 1, we showed that for an object moving along a straight line with position function $s(t)$, the object's "average velocity on the interval $[a, b]$ " is given by

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

More recently in Chapter 4, we found that for an object moving along a straight line with velocity function $v(t)$, the object's "average value of its velocity function on $[a, b]$ " is

$$v_{\text{AVG}}_{[a,b]} = \frac{1}{b - a} \int_a^b v(t) dt.$$

Are the "average velocity on the interval $[a, b]$ " and the "average value of the velocity function on $[a, b]$ " the same thing? Why or why not? Explain.

Answer. Yes.

Solution. Yes, $AV_{[a,b]} = v_{\text{AVG}}_{[a,b]}$. This is true because of the Fundamental Theorem of Calculus and the definitions of these two quantities. In particular,

$$\begin{aligned} v_{\text{AVG}}_{[a,b]} &= \frac{1}{b - a} \int_a^b v(t) dt \\ &= \frac{s(b) - s(a)}{b - a} \\ &= AV_{a,b}. \end{aligned}$$

Average velocity can be thought of in two very different ways: one as the change in position divided by change in time, the other as an average of the values of $v(t)$ itself.

Chapter 5

Evaluating Integrals

5.1 Constructing Accurate Graphs of Antiderivatives

Motivating Questions

- Given the graph of a function's derivative, how can we construct a completely accurate graph of the original function?
 - How many antiderivatives does a given function have? What do those antiderivatives all have in common?
 - Given a function f , how does the rule $A(x) = \int_0^x f(t) dt$ define a new function A ?
-

A recurring theme in our discussion of differential calculus has been the question “Given information about the derivative of an unknown function f , how much information can we obtain about f itself?” In Example 2.9.8, the graph of $y = f'(x)$ was known (along with the value of f at a single point) and we endeavored to sketch a possible graph of f near the known point. In Example 3.1.8 — we investigated how the first derivative test enables us to use information about f' to determine where the original function f is increasing and decreasing, as well as where f has relative extreme values. If we know a formula or graph of f' , by computing f'' we can find where the original function f is concave up and concave down. Thus, knowing f' and f'' enables us to understand the shape of the graph of f .

We returned to this question in even more detail in Section 4.1. In that setting, we knew the instantaneous velocity of a moving object and worked to determine as much as possible about the object's position function. We found connections between the net signed area under the velocity function and the corresponding change in position of the function, and the Total Change Theorem further illuminated these connections between f' and f , showing that the total change in the value of f over an interval $[a, b]$ is determined by the net signed

area bounded by f' and the x -axis on the same interval.

In what follows, we explore the situation where we possess an accurate graph of the derivative function along with a single value of the function f . From that information, we'd like to determine a graph of f that shows where f is increasing, decreasing, concave up, and concave down, and also provides an accurate function value at any point.

Example 5.1.1. Suppose that the following information is known about a function f : the graph of its derivative, $y = f'(x)$, is given in Figure 5.1.2. Further, assume that f' is piecewise linear (as pictured) and that for $x \leq 0$ and $x \geq 6$, $f'(x) = 0$. Finally, it is given that $f(0) = 1$.

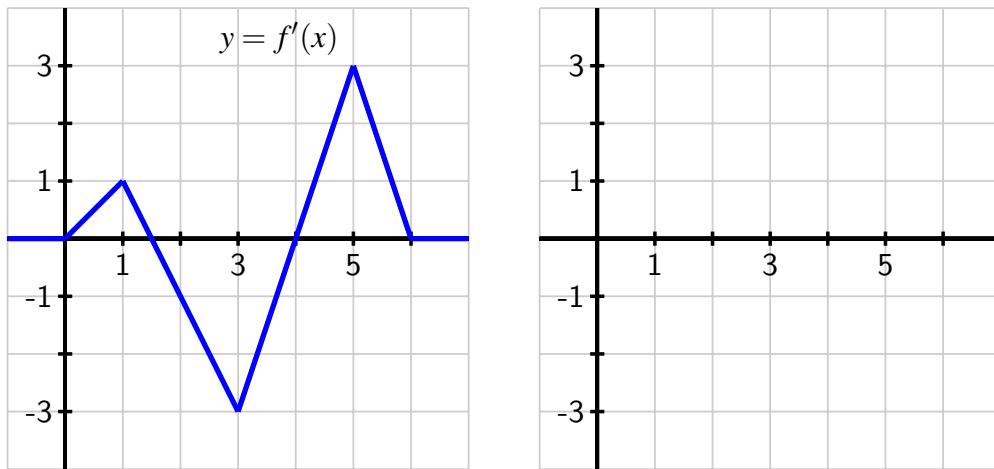


Figure 5.1.2: At left, the graph of $y = f'(x)$; at right, axes for plotting $y = f(x)$.

- On what interval(s) is f an increasing function? On what intervals is f decreasing?
- On what interval(s) is f concave up? concave down?
- At what point(s) does f have a relative minimum? a relative maximum?
- Recall that the Total Change Theorem tells us that

$$f(1) - f(0) = \int_0^1 f'(x) dx.$$

What is the exact value of $f(1)$?

- Use the given information and similar reasoning to that in (d) to determine the exact value of $f(2), f(3), f(4), f(5)$, and $f(6)$.
- Based on your responses to all of the preceding questions, sketch a complete and accurate graph of $y = f(x)$ on the axes provided, being sure to indicate the behavior of f for $x < 0$ and $x > 6$.

5.1.1 Constructing the graph of an antiderivative

Example 5.1.1 demonstrates that when we can find the exact area under the graph of a function on any given interval, it is possible to construct a graph of the function's antiderivative. That is, we can find a function whose derivative is given. We can now determine not only the overall shape of the antiderivative graph, but also the actual *height* of the graph at any point of interest.

This is a consequence of the Fundamental Theorem of Calculus: if we know a function f and the value of the antiderivative F at some starting point a , we can determine the value of $F(b)$ via the definite integral. Since $F(b) - F(a) = \int_a^b f(x) dx$, it follows that

$$F(b) = F(a) + \int_a^b f(x) dx. \quad (5.1.1)$$

We can also interpret the equation $F(b) - F(a) = \int_a^b f(x) dx$ in terms of the graphs of f and F as follows. On an interval $[a, b]$,

differences in heights on the graph of the antiderivative given by $F(b) - F(a)$ correspond to the net signed area bounded by the original function on the interval $[a, b]$, which is given by $\int_a^b f(x) dx$.

Example 5.1.3. Suppose that the function $y = f(x)$ is given by the graph shown in Figure 5.1.4, and that the pieces of f are either portions of lines or portions of circles. In addition, let F be an antiderivative of f and say that $F(0) = -1$. Finally, assume that for $x \leq 0$ and $x \geq 7$, $f(x) = 0$.

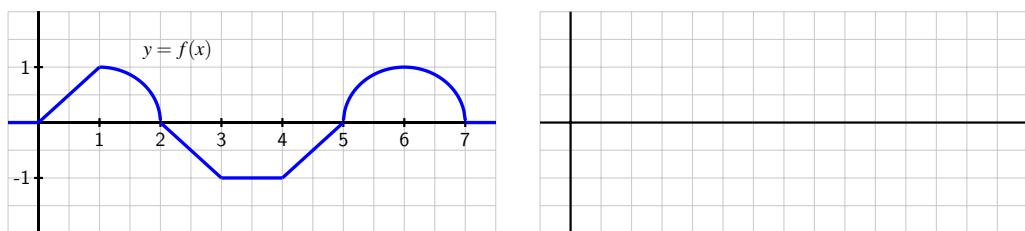


Figure 5.1.4: At left, the graph of $y = f(x)$.

- On what interval(s) is F an increasing function? On what intervals is F decreasing?
- On what interval(s) is F concave up? concave down? neither?
- At what point(s) does F have a relative minimum? a relative maximum?
- Use the given information to determine the exact value of $F(x)$ for $x = 1, 2, \dots, 7$. In addition, what are the values of $F(-1)$ and $F(8)$?

- e. Based on your responses to all of the preceding questions, sketch a complete and accurate graph of $y = F(x)$ on the axes provided, being sure to indicate the behavior of F for $x < 0$ and $x > 7$. Clearly indicate the scale on the vertical and horizontal axes of your graph.
- f. What happens if we change one key piece of information: in particular, say that G is an antiderivative of f and $G(0) = 0$. How (if at all) would your answers to the preceding questions change? Sketch a graph of G on the same axes as the graph of F you constructed in (e).

Hint.

- a. Consider the sign of $F' = f$.
- b. Consider the sign of $F'' = f'$.
- c. Where does $F' = f$ change sign?
- d. Recall that $F(1) = F(0) + \int_0^1 f(t) dt$.
- e. Use the function values found in (d) and the earlier information regarding the shape of F .
- f. Note that $G(1) = G(0) + \int_0^1 f(t) dt$.

Answer.

- a. F is increasing on $(0, 2)$ and $(5, 7)$; F is decreasing on $(2, 5)$.
- b. F is concave up on $(0, 1)$, $(4, 6)$; concave down on $(1, 3)$, $(6, 7)$; neither on $(3, 4)$.
- c. A relative maximum at $x = 2$; a relative minimum at $x = 5$.
- d. $F(1) = -\frac{1}{2}$; $F(2) = \frac{\pi}{4} - \frac{1}{2}$; $F(3) = \frac{\pi}{4} - 1$; $F(4) = \frac{\pi}{4} - 2$; $F(5) = \frac{\pi}{4} - \frac{5}{2}$; $F(6) = \frac{\pi}{2} - \frac{5}{2}$; $F(7) = \frac{3\pi}{4} - \frac{5}{2}$; $F(8) = \frac{3\pi}{4} - \frac{5}{2}$; and $F(-1) = -1$.
- e. Use the function values found in (d) and the earlier information regarding the shape of F .
- f. $G(x) = F(x) + 1$.

Solution.

- a. Wherever $F' > 0$, F is increasing, so F is increasing on $(0, 2)$ and $(5, 7)$, while F is decreasing on $(2, 5)$.
- b. Wherever $F'' > 0$, F is concave up; note particularly that $F'' > 0$ if and only if f is increasing. Thus, F is concave up on $(0, 1)$, $(4, 6)$, and concave down on $(1, 3)$, $(6, 7)$, and neither on $(3, 4)$.
- c. A relative maximum for F will occur wherever F' changes from positive to negative, and thus at $x = 2$; similarly, F has a relative minimum at $x = 5$.

- d. Recall that $F(1) = F(0) + \int_0^1 f(t) dt$, so $F(1) = -1 + \frac{1}{2} = -\frac{1}{2}$. Similarly, $F(2) = F(0) + \int_0^2 f(t) dt = -1 + \frac{1}{2} + \frac{\pi}{4} = \frac{\pi}{4} - \frac{1}{2}$. Continuing these calculations, $F(3) = \frac{\pi}{4} - 1$, $F(4) = \frac{\pi}{4} - 2$, $F(5) = \frac{\pi}{4} - \frac{5}{2}$, $F(6) = \frac{\pi}{2} - \frac{5}{2}$, $F(7) = \frac{3\pi}{4} - \frac{5}{2}$. Furthermore, since $f(t) = 0$ for all $t < 0$ and all $t > 7$, it follows $F(8) = \frac{3\pi}{4} - \frac{5}{2}$ and $F(-1) = -1$.
- e. Use the function values found in (d) and the earlier information regarding the shape of F .
- f. Note that $G(1) = G(0) + \int_0^1 f(t) dt$. Since $G(0) = 0$ (while $F(0) = -1$), this changes each response in by 1: $G(x) = F(x) + 1$.

5.1.2 Multiple antiderivatives of a single function

In the final question of Example 5.1.3, we encountered a very important idea: a function f has more than one antiderivative. Each antiderivative of f is determined uniquely by its value at a single point. For example, suppose that f is the function given at left in Figure 5.1.5, and suppose further that F is an antiderivative of f that satisfies $F(0) = 1$.

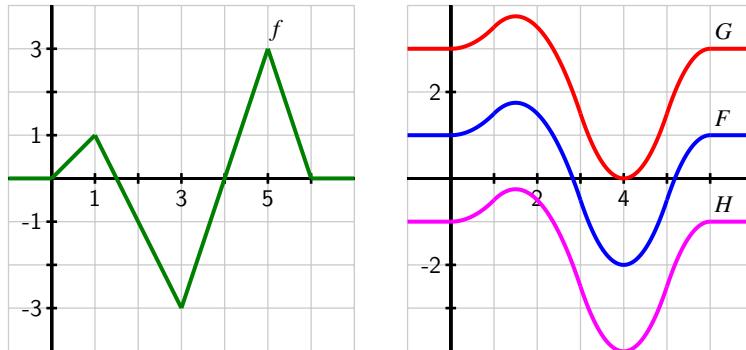


Figure 5.1.5: At left, the graph of $y = f(x)$. At right, three different antiderivatives of f .

Then, using Equation (5.1.1), we can compute

$$\begin{aligned} F(1) &= F(0) + \int_0^1 f(x) dx \\ &= 1 + 0.5 \\ &= 1.5. \end{aligned}$$

Similarly, $F(2) = 1.5$, $F(3) = -0.5$, $F(4) = -2$, $F(5) = -0.5$, and $F(6) = 1$. In addition, we can use the fact that $F' = f$ to ascertain where F is increasing and decreasing, concave up and concave down, and has relative extremes and inflection points. We ultimately find that the graph of F is the one given in blue in Figure 5.1.5.

If we want an antiderivative G for which $G(0) = 3$, then G will have the exact same shape as F (since both share the derivative f), but G will be shifted vertically from the graph of

F , as pictured in red in Figure 5.1.5. Note that $G(1) - G(0) = \int_0^1 f(x) dx = 0.5$, just as $F(1) - F(0) = 0.5$, but since $G(0) = 3$, $G(1) = G(0) + 0.5 = 3.5$, whereas $F(1) = 1.5$. In the same way, if we assigned a different initial value to the antiderivative, say $H(0) = -1$, we would get still another antiderivative, as shown in magenta in Figure 5.1.5.

This example demonstrates an important fact that holds more generally:

If G and H are both antiderivatives of a function f , then the function $G - H$ must be constant.

To see why this result holds, observe that if G and H are both antiderivatives of f , then $G' = f$ and $H' = f$. Hence,

$$\frac{d}{dx}[G(x) - H(x)] = G'(x) - H'(x) = f(x) - f(x) = 0.$$

Since the only way a function can have derivative zero is by being a constant function, it follows that the function $G - H$ must be constant.

We now see that if a function has at least one antiderivative, it must have infinitely many: we can add any constant of our choice to the antiderivative and get another antiderivative. For this reason, we sometimes refer to the *general antiderivative* of a function f .

To identify a particular antiderivative of f , we must know a single value of the antiderivative F (this value is often called an *initial condition*). For example, if $f(x) = x^2$, its general antiderivative is $F(x) = \frac{1}{3}x^3 + C$, where we include the “ $+C$ ” to indicate that F includes *all* of the possible antiderivatives of f . If we know that $F(2) = 3$, we substitute 2 for x in $F(x) = \frac{1}{3}x^3 + C$, and find that

$$3 = \frac{1}{3}(2)^3 + C,$$

or $C = 3 - \frac{8}{3} = \frac{1}{3}$. Therefore, the particular antiderivative in this case is $F(x) = \frac{1}{3}x^3 + \frac{1}{3}$.

Example 5.1.6. For each of the following functions, sketch an accurate graph of the antiderivative that satisfies the given initial condition. In addition, sketch the graph of two additional antiderivatives of the given function, and state the corresponding initial conditions that each of them satisfy. If possible, find an algebraic formula for the antiderivative that satisfies the initial condition.

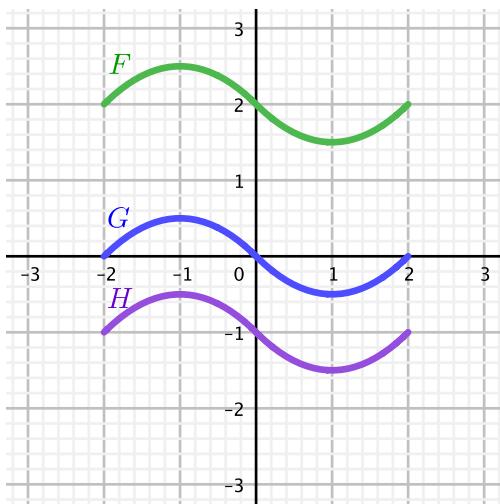
a. original function: $g(x) = |x| - 1$; initial condition: $G(-1) = 0$; interval for sketch: $[-2, 2]$

b. original function: $h(x) = \sin(x)$; initial condition: $H(0) = 1$; interval for sketch: $[0, 4\pi]$

c. original function: $p(x) = \begin{cases} x^2, & \text{if } 0 < x \leq 1 \\ -(x-2)^2, & \text{if } 1 < x < 2; \text{ initial condition: } P(0) = 1; \text{ interval for sketch: } [-1, 3] \\ 0 & \text{otherwise} \end{cases}$

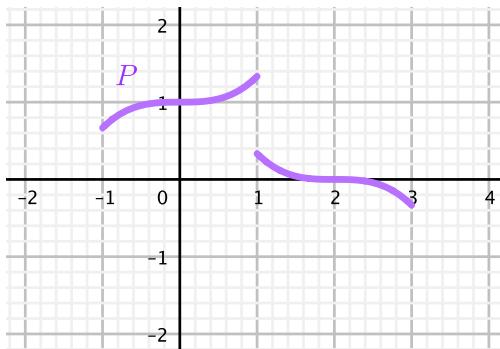
Answer.

a.



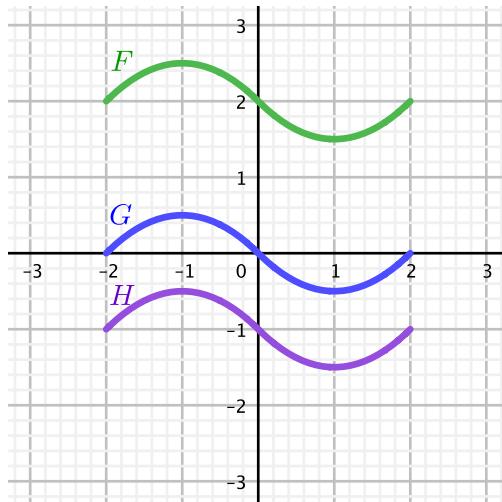
b. $H(x) = -\cos(x) + 2$.

c.

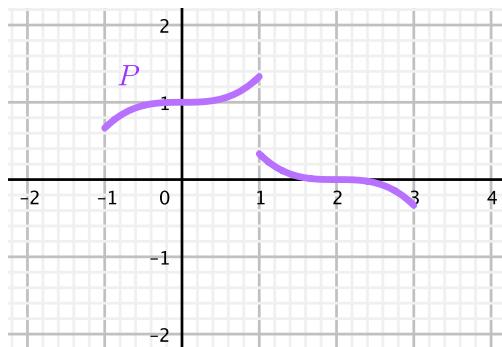


Solution.

- a. A possible antiderivative G that satisfies $G(0) = 0$ is shown in the figure below, with a possible formula being $G(x) = -\frac{1}{2}x^2 - x$ for $x < 0$ and $G(x) = \frac{1}{2}x^2 - x$ for $x \geq 0$. Other antiderivatives that satisfy $F(0) = 2$ and $H(0) = -1$ are also shown.



- b. Thinking both graphically and algebraically reveals that the antiderivative we seek is $H(x) = -\cos(x) + 2$. Any vertical shift of this function will have the same derivative, but will have a different y -intercept.
- c. A possible antiderivative P that satisfies $P(0) = 1$ is shown in the figure below.



5.1.3 Functions defined by integrals

Equation (5.1.1) allows us to compute the value of the antiderivative F at a point b , provided that we know $F(a)$ and can evaluate the definite integral from a to b of f . That is,

$$F(b) = F(a) + \int_a^b f(x) dx.$$

In several situations, we have used this formula to compute $F(b)$ for several different values of b , and then plotted the points $(b, F(b))$ to help us draw an accurate graph of F . This suggests that we may want to think of b , the upper limit of integration, as a variable itself. To that end, we introduce the idea of an *integral function*, a function whose formula involves a definite integral.

Definition 5.1.7. If f is a continuous function, we define the corresponding **integral function** A according to the rule

$$A(x) = \int_a^x f(t) dt. \quad (5.1.2)$$

Note that because x is the independent variable in the function A , and determines the endpoint of the interval of integration, we need to use a different variable 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 $y = f(t)$. For example, in Figure 5.1.8, we see a function f pictured at left, and its corresponding area function (choosing $a = 0$), $A(x) = \int_0^x f(t) dt$ shown at right.

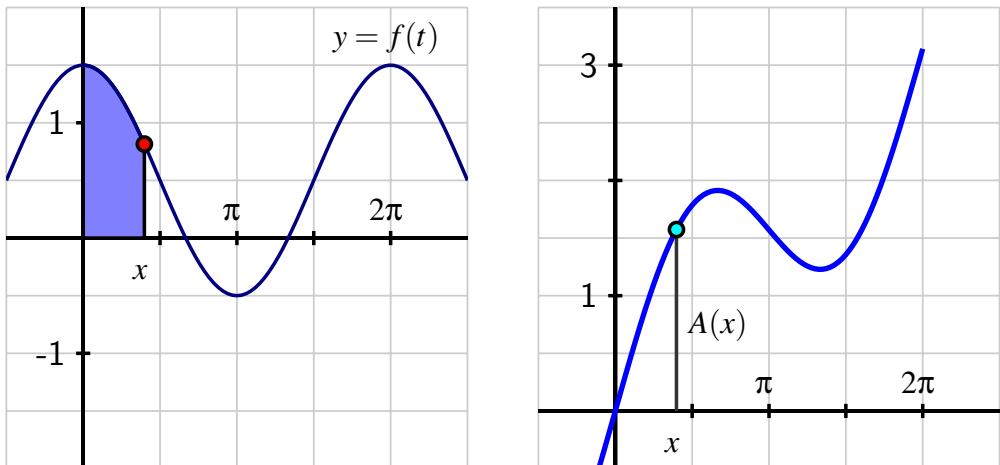


Figure 5.1.8: At left, the graph of the given function f . At right, the area function $A(x) = \int_0^x f(t) dt$.

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 $y = A(x)$. At <http://gvsu.edu/s/cz>, we find a java applet¹ that brings the static picture in Figure 5.1.8 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 .

The choice of a is somewhat arbitrary. In the example that follows, we explore how the value of a affects the graph of the integral function.

Example 5.1.9. Suppose that g is given by the graph at left in Figure 5.1.10 and that A is the corresponding integral function defined by $A(x) = \int_1^x g(t) dt$.

¹David Austin, Grand Valley State University

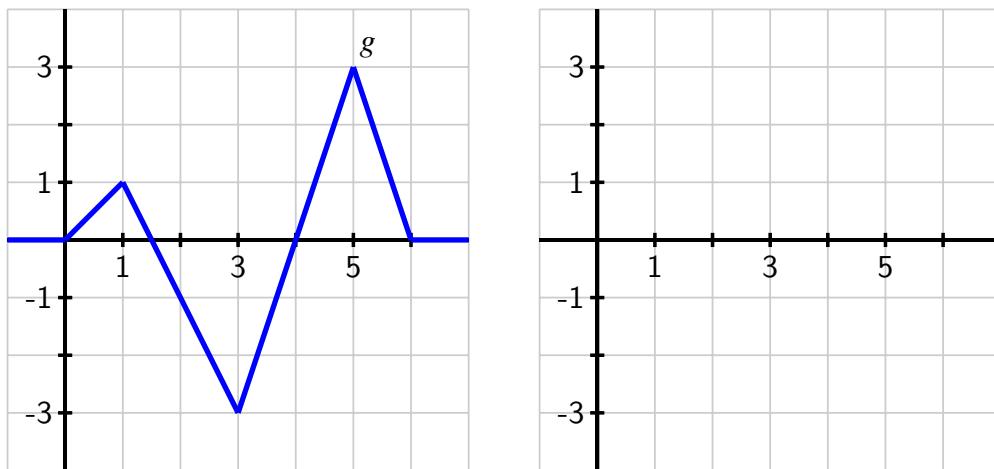


Figure 5.1.10: At left, the graph of $y = g(t)$; at right, axes for plotting $y = A(x)$, where A is defined by the formula $A(x) = \int_1^x g(t) dt$.

- On what interval(s) is A an increasing function? On what intervals is A decreasing? Why?
- On what interval(s) do you think A is concave up? concave down? Why?
- At what point(s) does A have a relative minimum? a relative maximum?
- Use the given information to determine the exact values of $A(0)$, $A(1)$, $A(2)$, $A(3)$, $A(4)$, $A(5)$, and $A(6)$.
- 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 $x > 6$.
- How does the graph of B compare to A if B is instead defined by $B(x) = \int_0^x g(t) dt$?

Hint.

- Where is A accumulating positive signed area?
- As A accumulates positive or negative signed area, where is the rate at which such area is accumulated increasing?
- Where does A change from accumulating positive signed area to accumulating negative signed area?
- Note, for instance, that $A(2) = \int_1^2 g(t) dt$.

- e. Use your work in (a)-(d) appropriately.
- f. What is the value of $B(0)$? How does this compare to $A(0)$?

Answer.

- a. A is increasing on $(0, 1.5), (4, 6)$; A is decreasing on $(1.5, 4)$.
- b. A is concave up on $(0, 1)$ and $(3, 5)$; A is concave down on $(1, 3)$ and $(5, 6)$.
- c. At $x = 1.5$, A has a relative maximum; A has a relative minimum at $x = 4$.
- d. $A(0) = -\frac{1}{2}$; $A(1) = 0$; $A(2) = 0$; $A(3) = -2$; $A(4) = -3.5$, $A(5) = -2$, $A(6) = -0.5$.
- e. Use your work in (a)-(d) appropriately.
- f. $B(x) = A(x) + \frac{1}{2}$.

Solution.

- a. A is accumulating positive signed area wherever g is positive, and thus A is increasing on $(0, 1.5), (4, 6)$; A is accumulating negative signed area and therefore decreasing wherever g is negative, which occurs on $(1.5, 4)$.
- b. Here we want to consider where A is changing at an increasing rate (concave up) or changing at a decreasing rate (concave down). On $(0, 1)$ and $(4, 5)$, A is increasing, and we can also see that since g is increasing, A is increasing at an increasing rate. Similarly, on $(3, 4)$ (where g is negative so A is decreasing), since g is increasing it follows that A is decreasing at an increasing rate. Thus, A is concave up on $(0, 1)$ and $(3, 5)$. Analogous reasoning shows that A is concave down on $(1, 3)$ and $(5, 6)$.
- c. Based on our work in (a), we see that A changes from increasing to decreasing at $x = 1.5$, and thus A has a relative maximum there. Similarly, A has a relative minimum at $x = 4$.
- d. Using the fact that g is piecewise linear and the definition of A , we find that $A(0) = \int_1^0 g(t) dt = -\int_0^1 g(t) dt = -\frac{1}{2}$; $A(1) = \int_1^1 g(t) dt = 0$; $A(2) = \int_1^2 g(t) dt = 0$. Analogous reasoning shows that $A(3) = -2$, $A(4) = -3.5$, $A(5) = -2$, $A(6) = -0.5$.
- e. Use your work in (a)-(d) appropriately.
- f. Note that $B(0) = 0$, while $A(0) = -\frac{1}{2}$. Likewise, $B(1) = \frac{1}{2}$, while $A(1) = 0$. Indeed, we can see that for any value of x , $B(x) = A(x) + \frac{1}{2}$.

5.1.4 Summary

- Given the graph of a function f , we can construct the graph of its antiderivative F provided that (a) we know a starting value of F , say $F(a)$, and (b) we can evaluate the integral $\int_a^b f(x) dx$ exactly for relevant choices of a and b . For instance, if we wish to

know $F(3)$, we can compute $F(3) = F(a) + \int_a^3 f(x) dx$. When we combine this information about the function values of F together with our understanding of how the behavior of $F' = f$ affects the overall shape of F , we can develop a completely accurate graph of the antiderivative F .

- Because the derivative of a constant is zero, if F is an antiderivative of f , it follows that $G(x) = F(x) + C$ will also be an antiderivative of f . Moreover, any two antiderivatives of a function f differ precisely by a constant. Thus, any function with at least one antiderivative in fact has infinitely many, and the graphs of any two antiderivatives will differ only by a vertical translation.
- 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 .

5.1.5 Exercises

1. (Definite integral of a piecewise linear function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (A smooth function that starts out at 0) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (A piecewise constant function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Another piecewise linear function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. A moving particle has its velocity given by the quadratic function v pictured in Figure 5.1.11. In addition, it is given that $A_1 = \frac{7}{6}$ and $A_2 = \frac{8}{3}$, as well as that for the corresponding position function s , $s(0) = 0.5$.
 - a. Use the given information to determine $s(1)$, $s(3)$, $s(5)$, and $s(6)$.
 - b. On what interval(s) is s increasing? On what interval(s) is s decreasing?
 - c. On what interval(s) is s concave up? On what interval(s) is s concave down?
 - d. Sketch an accurate, labeled graph of s on the axes at right in Figure 5.1.11.
 - e. Note that $v(t) = -2 + \frac{1}{2}(t - 3)^2$. Find a formula for s .

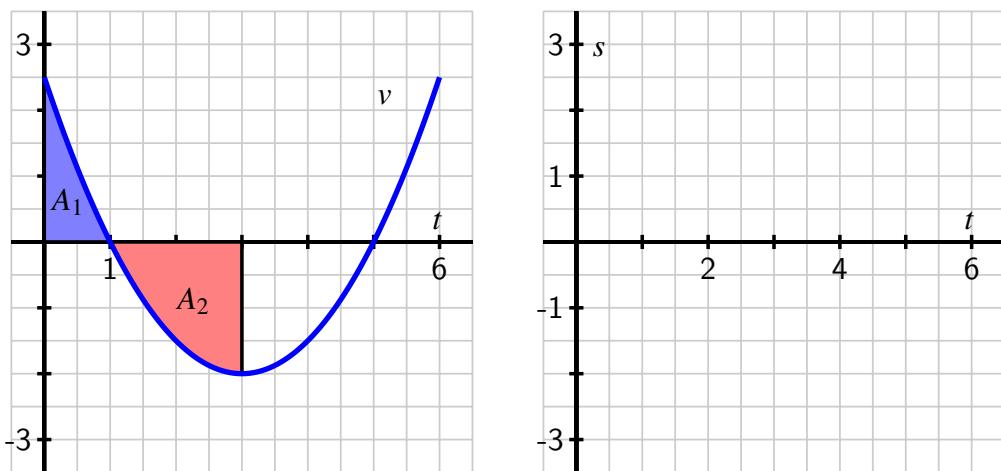
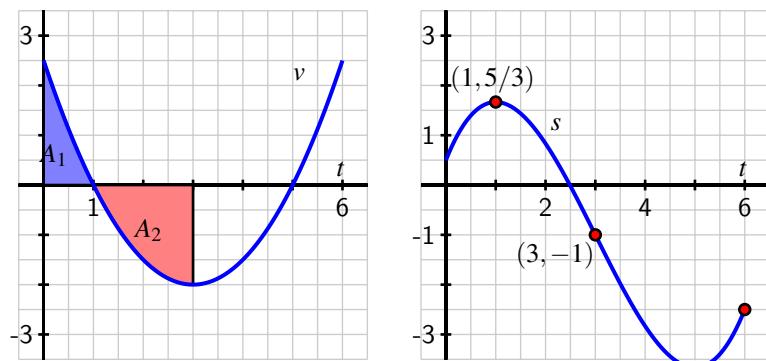


Figure 5.1.11: At left, the given graph of v . At right, axes for plotting s .

Answer.

- $s(1) = \frac{5}{3}$, $s(3) = -1$, $s(5) = -\frac{11}{3}$, $s(6) = -\frac{5}{2}$.
- s is increasing on $0 < t < 1$ and $5 < t < 6$; decreasing for $1 < t < 5$.
- s is concave down for $t < 2$; concave up for $t > 2$.
-



- $s(t) = -2t + \frac{1}{6}(t-3)^3 + 5$.

Solution.

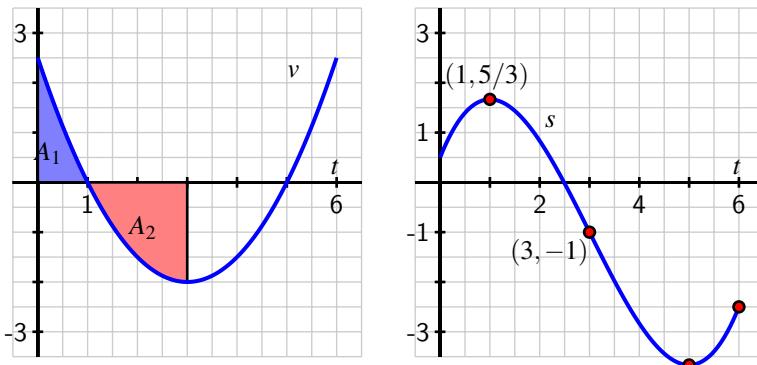
- We use the Fundamental Theorem of Calculus repeatedly together with the given information and the symmetry of the function v , while paying careful attention to signed area:

- $s(1) = s(0) + \int_0^1 v(t) dt = \frac{1}{2} + \frac{7}{6} = \frac{5}{3}$
- $s(3) = s(1) + \int_1^3 v(t) dt = \frac{5}{3} - \frac{8}{3} = -1$

- $s(5) = s(3) + \int_3^5 v(t) dt = -1 - \frac{8}{3} = -\frac{11}{3}$

- $s(6) = s(5) + \int_5^6 v(t) dt = -\frac{11}{3} + \frac{7}{6} = -\frac{5}{2}$

- b. The position function s is increasing on $0 < t < 1$ and $5 < t < 6$ since this is where $v(t) = s'(t)$ is positive. Similarly, s is decreasing for $1 < t < 5$.
- c. The position function s is concave down for $t < 2$ since $v = s'$ is decreasing on this interval, and s is concave up for $t > 2$ because $v = s'$ increases there.
- d. In the following figure, we see a plot of s that reflects all of the known information, including the points $(1, \frac{5}{3})$, $(3, -1)$, $(5, -\frac{11}{3})$, and $(6, -2.5)$.



- e. The general antiderivative of $v(t) = -2 + \frac{1}{2}(t-3)^2$ is $s(t) = -2t + \frac{1}{6}(t-3)^3 + C$. Given that $s(0) = \frac{1}{2}$, we know that $\frac{1}{2} = \frac{1}{6}(-3)^3 + C$, and thus $C = \frac{1}{2} + \frac{27}{6} = \frac{30}{6} = 5$. Hence, $s(t) = -2t + \frac{1}{6}(t-3)^3 + 5$.
6. A person exercising on a treadmill experiences different levels of resistance and thus burns calories at different rates, depending on the treadmill's setting. In a particular workout, the rate at which a person is burning calories is given by the piecewise constant function c pictured in Figure 5.1.12. Note that the units on c are "calories per minute."

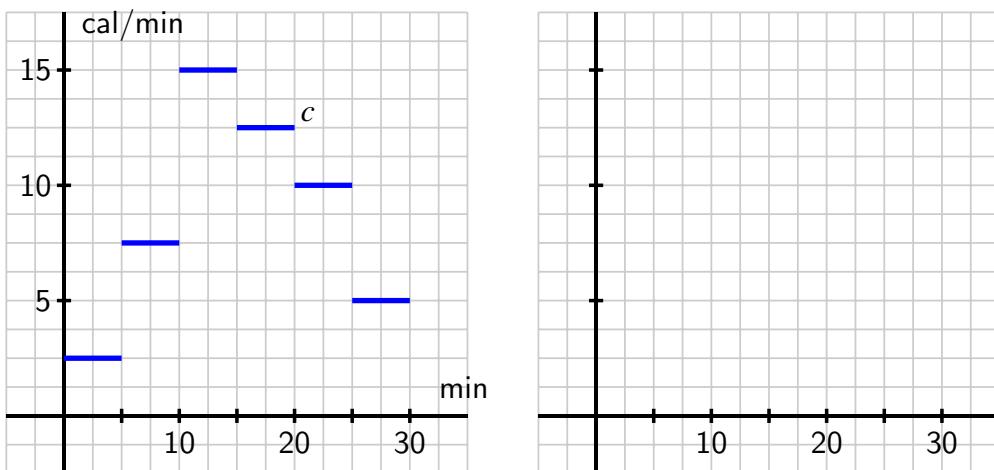
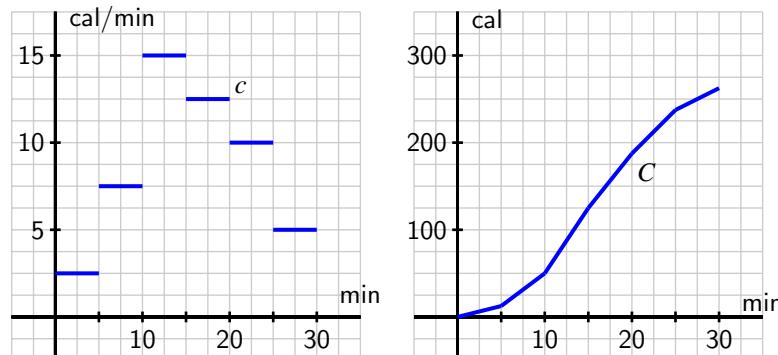


Figure 5.1.12: At left, the given graph of c . At right, axes for plotting C .

- Let C be an antiderivative of c . What does the function C measure? What are its units?
- Assume that $C(0) = 0$. Determine the exact value of $C(t)$ at the values $t = 5, 10, 15, 20, 25, 30$.
- Sketch an accurate graph of C on the axes provided at right in Figure 5.1.12. Be certain to label the scale on the vertical axis.
- Determine a formula for C that does not involve an integral and is valid for $5 \leq t \leq 10$.

Answer.

- C measures the total number of calories burned in the workout since $t = 0$.
- $C(5) = 12.5, C(10) = 50, C(15) = 125, C(20) = 187.5, C(25) = 237.5, C(30) = 262.5$.
- c.

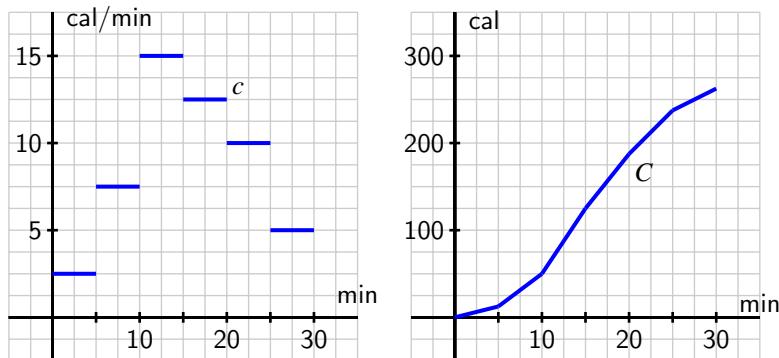


- $C(t) = 12.5 + 7.5(t - 5)$ on this interval.

Solution.

- a. An antiderivative function C measures the total number of calories burned in the workout since $t = 0$. Since the units on $c(t) = C'(t)$ are “calories per minute”, the units on $C(t)$ are “calories”.
- b. The value of $C(5) = C(0) + \int_0^5 c(t) dt$. Since c is piecewise constant, the value of the integral is simply the area of the rectangle bounded by c , so $C(5) = C(0) + (2.5)(5) = 0 + 12.5 = 12.5$. Reasoning similarly,
- $C(10) = C(5) + \int_5^{10} c(t) dt = 12.5 + (7.5)(5) = 50$
 - $C(15) = C(10) + \int_{10}^{15} c(t) dt = 50 + (15)(5) = 125$
 - $C(20) = C(15) + \int_{15}^{20} c(t) dt = 125 + (12.5)(5) = 187.5$
 - $C(25) = C(20) + \int_{20}^{25} c(t) dt = 187.5 + (10)(5) = 237.5$
 - $C(30) = C(25) + \int_{25}^{30} c(t) dt = 237.5 + (5)(5) = 262.5$

- c. In the figure below, we see that C is a piecewise linear function that passes through the points we established in (b) (such as $(15, 125)$ and $(20, 187.5)$) and that the slope of C on each interval where c is constant is simply the value of $c(t)$ on that interval (such as $m = 12.5$ on $15 < t < 20$).



- d. C is linear on $5 \leq t \leq 10$ with slope $m = 7.5$ and passes through the point $(5, C(5)) = (5, 12.5)$. Thus, $C(t) = 12.5 + 7.5(t - 5)$ on this interval.
7. Consider the piecewise linear function f given in Figure 5.1.13. Let the functions A , B , and C be defined by the rules $A(x) = \int_{-1}^x f(t) dt$, $B(x) = \int_0^x f(t) dt$, and $C(x) = \int_1^x f(t) dt$.
- a. For the values $x = -1, 0, 1, \dots, 6$, make a table that lists corresponding values of $A(x)$, $B(x)$, and $C(x)$.
- b. On the axes provided in Figure 5.1.13, sketch the graphs of A , B , and C .
- c. How are the graphs of A , B , and C related?
- d. How would you best describe the relationship between the function A and the function f ?

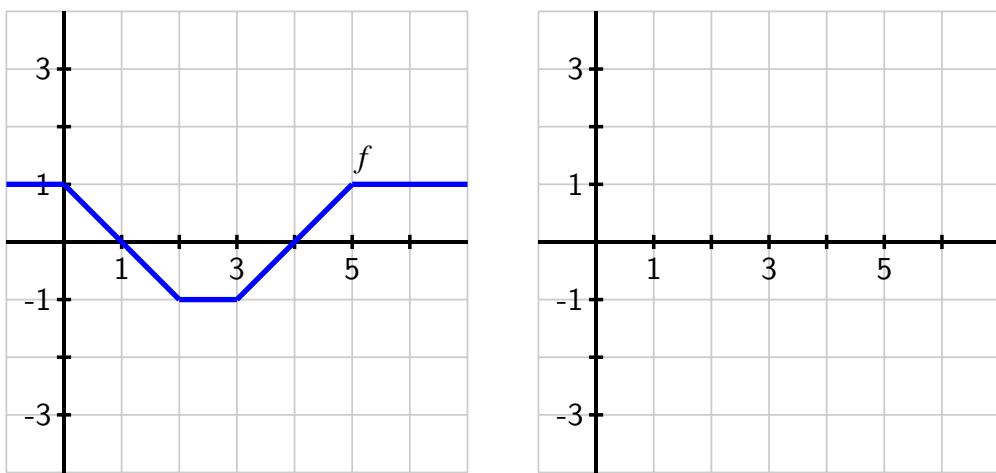


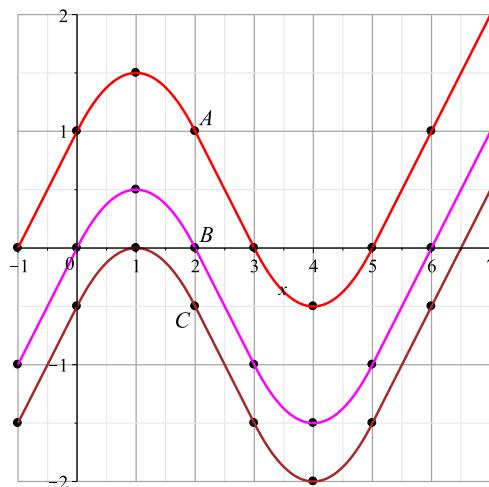
Figure 5.1.13: At left, the given graph of f . At right, axes for plotting A , B , and C .

Answer.

- a. $B(-1) = -1$, $B(0) = 0$, $B(1) = \frac{1}{2}$, $B(2) = 0$, $B(3) = -1$, $B(4) = -\frac{3}{2}$, $B(5) = -1$, $B(6) = 0$.
 Also, $A(x) = 1 + B(x)$ and $C(x) = B(x) - \frac{1}{2}$.

x	-1	0	1	2	3	4	5	6
$A(x)$	0	1	1.5	1	0	-0.5	0	1
$B(x)$	-1	0	0.5	0	-1	-1.5	-1	0
$C(x)$	-1.5	-0.5	0	-0.5	-1.5	-2	-1.5	-0.5

b.



- c. A , B , and C are vertical translations of each other.

d. $A' = f$.

Solution.

- a. We use the First Fundamental Theorem and areas under the curves to complete the desired table. As an example, first note that

$$B(0) = \int_0^0 f(t) dt = 0.$$

Then, since the net-signed area between the graph of f and the x -axis on the interval $[-1, 0]$ is the area of a square with side lengths of 1, we have

$$B(0) - B(-1) = \int_{-1}^0 f(t) dt$$

and thus $B(0) - B(-1) = 1$, so $0 - B(-1) = 1$ and $B(-1) = -1$.

$$\begin{aligned} B(1) &= B(0) + \int_0^1 f(t) dt = 0 + \frac{1}{2} = \frac{1}{2}, \\ B(2) &= B(1) + \int_1^2 f(t) dt = \frac{1}{2} + \left(-\frac{1}{2}\right) = 0, \\ B(3) &= B(2) + \int_2^3 f(t) dt = 0 + (-1) = -1, \\ B(4) &= B(3) + \int_3^4 f(t) dt = -1 + \left(-\frac{1}{2}\right) = -\frac{3}{2}, \\ B(5) &= B(4) + \int_4^5 f(t) dt = -\frac{3}{2} + \frac{1}{2} = -1, \\ B(6) &= B(5) + \int_5^6 f(t) dt = -1 + 1 = 0. \end{aligned}$$

We can also say that

$$A(x) = \int_{-1}^x f(t) dt = \int_{-1}^0 f(t) dt + \int_0^x f(t) dt = 1 + B(x)$$

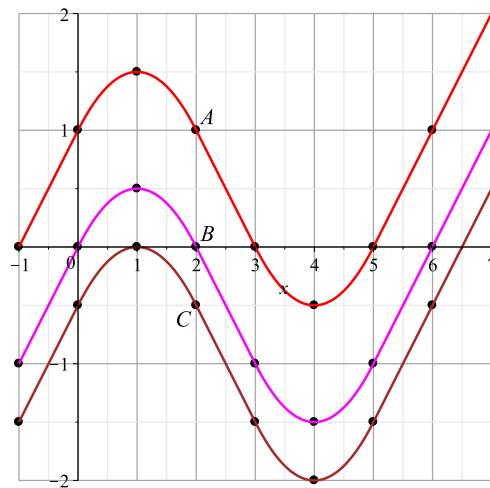
and

$$C(x) = \int_1^x f(t) dt = \int_0^x f(t) dt - \int_0^1 f(t) dt = B(x) - \frac{1}{2}.$$

This allows us to complete the following table.

x	-1	0	1	2	3	4	5	6
$A(x)$	0	1	1.5	1	0	-0.5	0	1
$B(x)$	-1	0	0.5	0	-1	-1.5	-1	0
$C(x)$	-1.5	-0.5	0	-0.5	-1.5	-2	-1.5	-0.5

- b. The graphs are shown in the figure below. Note that A , B , and C are all differentiable functions (that are antiderivatives of f , so each of their derivatives is f), so the graphs of these functions should be (and are) smooth.



- c. Notice that the graphs of A , B , and C are vertical translations of each other. More specifically, $A(x) = B(x) + 1$ and $B(x) = C(x) + 0.5$.
- d. Since A is an antiderivative of f , we can say that $A' = f$.

5.2 Antiderivatives from Formulas

Motivating Questions

- If we have a function given by a formula, is there a way to write down the antiderivative?
- How can we find the area between two curves?

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 the functions

$$5\sin(x) - 4x^2, x^2\sin(x), \frac{\sin(x)}{x^2}, \text{ and } \sin(x^2).$$

In the above cases, the antiderivative might not be obvious and in some cases, for example, $\sin(x^2)$ the antiderivative cannot be written down using elementary functions. This warrants the question, how do we find an antiderivative from a function given by a formula in general? We will begin to answer that question in this chapter.

5.2.1 Basic antiderivatives

What is involved in trying to find an antiderivative for each? From our experience with derivative rules, we know that derivatives of sums and constant multiples of basic functions are simple to execute, but derivatives involving products, quotients, and composites of familiar functions are more complicated. Therefore, 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.

Before proceeding any further it will be useful to state two properties.

Properties of Antiderivatives: Sums and Constant Multiples

Let f and g be functions that have an antiderivative and c a constant then

$$\int (f(x) \pm g(x)) dx = \int f(x)dx \pm \int g(x)dx$$

$$\int cf(x)dx = c \int f(x)dx.$$

In other words, the antiderivative of a sum is the sum of the antiderivatives.

We forgo a formal proof of these properties but note that they are a consequence of the Constant Multiple Rule and Sum Rule for derivatives in Section 2.1.

We do note that whenever we know the derivative of a function, we have a *function-derivative pair*, so we also know the antiderivative of a function. For instance, since we know that

$$\frac{d}{dx}[-\cos(x)] = \sin(x),$$

we also know that $F(x) = -\cos(x)$ is an antiderivative of $f(x) = \sin(x)$. F and f together form a function-derivative pair. Clearly, every basic derivative rule leads us to such a pair, and thus to a known antiderivative.

In Example 5.2.1, we will construct a list of the basic antiderivatives we know at this time. Those rules will help us antidifferentiate sums and constant multiples of basic functions. For example, since $-\cos(x)$ is an antiderivative of $\sin(x)$ and $\frac{1}{3}x^3$ is an antiderivative of x^2 , it follows that

$$F(x) = -5\cos(x) - \frac{4}{3}x^3$$

is an antiderivative of $f(x) = 5\sin(x) - 4x^2$.

Finally, before proceeding to build a list of common functions whose antiderivatives we know, we recall that each function has more than one antiderivative. Because the derivative of any constant is zero, we may add a constant of our choice to any antiderivative. For instance, we know that $G(x) = \frac{1}{3}x^3$ is an antiderivative of $g(x) = x^2$. But we could also have chosen $G(x) = \frac{1}{3}x^3 + 7$, since in this case as well, $G'(x) = x^2$. If $g(x) = x^2$, we say that the *general antiderivative* 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.

In the following example, we work to build a list of basic functions whose antiderivatives we already know.

Example 5.2.1. Use your knowledge of derivatives of basic functions to complete the above table of antiderivatives. For each entry, your task is to find a function F whose derivative is the given function f .

given function, $f(x)$	antiderivative, $F(x)$
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 5.2.2: Familiar basic functions and their antiderivatives.

Hint. You might start by constructing a list of all the basic functions whose derivative you know.

Answer.

given function, $f(x)$	antiderivative, $F(x)$
k , ($k \neq 0$)	$kx + C$
x^n , $n \neq -1$	$\frac{1}{n+1}x^{n+1} + C$
$\frac{1}{x}$, $x > 0$	$\ln(x) + C$
$\sin(x)$	$-\cos(x) + C$
$\cos(x)$	$\sin(x) + C$
$\sec(x) \tan(x)$	$\sec(x) + C$
$\csc(x) \cot(x)$	$-\csc(x) + C$
$\sec^2(x)$	$\tan(x) + C$
$\csc^2(x)$	$\cot(x) + C$
e^x	$e^x + C$
a^x ($a > 1$)	$\frac{1}{\ln(a)}a^x + C$
$\frac{1}{1+x^2}$	$\arctan(x) + C$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x) + C$

Solution.

given function, $f(x)$	antiderivative, $F(x)$
$k, (k \neq 0)$	$kx + C$
$x^n, n \neq -1$	$\frac{1}{n+1}x^{n+1} + C$
$\frac{1}{x}, x > 0$	$\ln(x) + C$
$\sin(x)$	$-\cos(x) + C$
$\cos(x)$	$\sin(x) + C$
$\sec(x) \tan(x)$	$\sec(x) + C$
$\csc(x) \cot(x)$	$-\csc(x) + C$
$\sec^2(x)$	$\tan(x) + C$
$\csc^2(x)$	$\cot(x) + C$
e^x	$e^x + C$
$a^x (a > 1)$	$\frac{1}{\ln(a)}a^x + C$
$\frac{1}{1+x^2}$	$\arctan(x) + C$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x) + C$

5.2.2 Evaluating Definite Integrals Involving Expressions

Now that we have some basic antiderivatives in hand we can proceed to solve problems involving the definite integral when the integrand is an expression. Specifically, we seek to use antiderivatives and the Fundamental Theorem of Calculus to evaluate expressions like

$$\int_0^1 (x^3 - x - e^x + 2) \, dx.$$

Since our current interest in antiderivatives is so that we can evaluate definite integrals by the Fundamental Theorem of Calculus the constant C is actually irrelevant, and mathematicians generally omit it in these types of problems. To see why, consider the definite integral

$$\int_0^1 x^2 \, dx.$$

For the integrand $g(x) = x^2$, suppose we find and use the general antiderivative $G(x) = \frac{1}{3}x^3 + C$. Then, by the FTC,

$$\begin{aligned} \int_0^1 x^2 \, dx &= \frac{1}{3}x^3 + C \Big|_0^1 \\ &= \left(\frac{1}{3}(1)^3 + C \right) - \left(\frac{1}{3}(0)^3 + C \right) \\ &= \frac{1}{3} + C - 0 - C \\ &= \frac{1}{3}. \end{aligned}$$

Observe that the C -values appear as opposites in the evaluation of the integral and thus do not affect the definite integral's value.

Example 5.2.3. Use the FTC and the results in Table 5.2.2 to evaluate the three given definite integrals.

a. $\int_0^1 (x^3 - x - e^x + 2) dx$

b. $\int_0^{\pi/3} (2 \sin(t) - 4 \cos(t) + \sec^2(t) - \pi) dt$

c. $\int_0^1 (\sqrt{x} - x^2) dx$

Hint. Be sure to recall the sum and constant multiple rules, which work not only for differentiating, but also for antiderivatives.

Answer.

a. $\int_0^1 (x^3 - x - e^x + 2) dx = \frac{11}{4} - 3.$

b. $\int_0^{\pi/3} (2 \sin(t) - 4 \cos(t) + \sec^2(t) - \pi) dt = 1 - \sqrt{3} - \frac{\pi^2}{3}.$

c. $\int_0^1 (\sqrt{x} - x^2) dx = \frac{1}{3}.$

Solution.

a. By standard antiderivative rules and the FTC,

$$\begin{aligned} \int_0^1 (x^3 - x - e^x + 2) dx &= \frac{1}{4}x^4 - \frac{1}{2}x^2 - e^x + 2x \Big|_0^1 \\ &= \left(\frac{1}{4}(1)^4 - \frac{1}{2}(1)^2 - e^1 + 2(1) \right) - \left(\frac{1}{4}(0)^4 - \frac{1}{2}(0)^2 - e^0 + 2(0) \right) \\ &= \frac{1}{4} - \frac{1}{2} - e + 2 - (0 - 0 - 1 + 0) \\ &= -\frac{1}{4} - e + 3 \\ &= \frac{11}{4} - 3. \end{aligned}$$

b. Calculating the needed antiderivative and applying the FTC,

$$\begin{aligned} \int_0^{\pi/3} (2 \sin(t) - 4 \cos(t) + \sec^2(t) - \pi) dt &= (-2 \cos(t) - 4 \sin(t) + \tan(t) - \pi t) \Big|_0^{\pi/3} \\ &= (-2 \cos(\pi/3) - 4 \sin(\pi/3) + \tan(\pi/3) - \pi(\pi/3)) - \\ &\quad (-2 \cos(0) - 4 \sin(0) + \tan(0) - \pi(0)) \\ &= -2 \cdot \frac{1}{2} - 4 \cdot \frac{\sqrt{3}}{2} + \sqrt{3} - \frac{\pi^2}{3} - (-2 - 0 + 0 - 0) \\ &= 1 - \sqrt{3} - \frac{\pi^2}{3}. \end{aligned}$$

c. Noting that $\frac{d}{dx}[\frac{2}{3}x^{3/2}] = x^{1/2}$, we find that

$$\begin{aligned}\int_0^1 (\sqrt{x} - x^2) dx &= \frac{2}{3}x^{3/2} - \frac{1}{3}x^3 \Big|_0^1 \\ &= \frac{2}{3} - \frac{1}{3} - (0 - 0) \\ &= \frac{1}{3}.\end{aligned}$$

5.2.3 Integration and the Area between Two Curves

Integration allows us to find the area between a curve and the x -axis. In fact, by visually inspecting the graphs of two different curves we can develop a technique for finding the area between two curves. Consider the graphs of the curves $y = x^2 + 1$ and $y = x + 3$ displayed in Figure 5.2.4. We might ask if there is a way to find the area enclosed by these two curves as is colored in the figure.

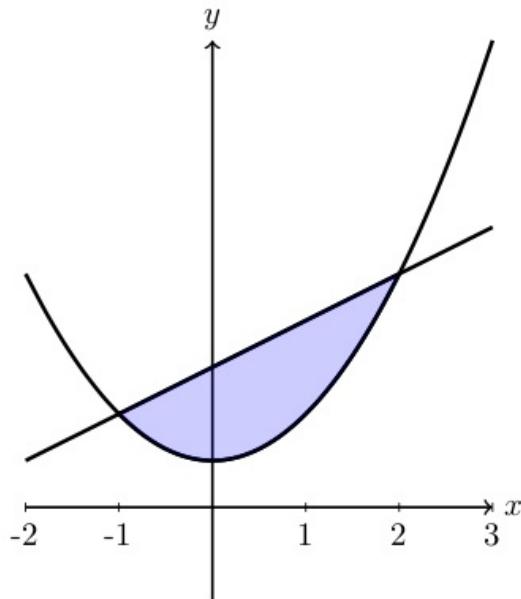


Figure 5.2.4: Graphs of the curves $y = x^2 + 1$ and $y = x + 3$.

In fact, using the knowledge we have about integration, we can figure out this quantity exactly.

Example 5.2.5. $f(x) = x^2 + 1$ $g(x) = x + 3$

Hint. It might help to answer the following questions:

- Where do the curves intersect?

- Based upon Figure 5.2.4 how would you calculate exact area between $g(x)$ and the x -axis between the points of intersection?
- Based upon Figure 5.2.4 how would you calculate exact area between $f(x)$ and the x -axis between the points of intersection?
- What is the relationship between the areas you calculated above?

Solution. The first step in solving this problem is to determine where the curves intersect. We can find the point of intersection by setting the two curves equal to each other.

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

Therefore, the curves intersect at two x -values, $x = -1, 2$. This means we are concerned with the interval $[-2, 2]$.

We next note that on the interval in question, $[-2, 2]$, the curve $g(x) = x + 3$ is above the curve $f(x) = x^2 + 1$. We also note that $\int_{-1}^2 (x + 3)dx$ would give the area shown in Figure 5.2.6 and $\int_{-1}^2 (x^2 + 1)dx$ Figure 5.2.7.

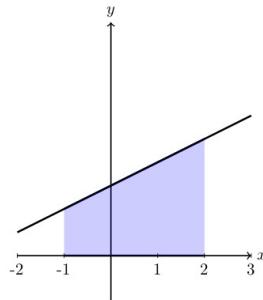


Figure 5.2.6: Graphs of the curves $y = x^2 + 1$ and $y = x + 3$.

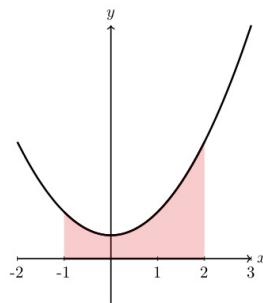


Figure 5.2.7: Graphs of the curves $y = x^2 + 1$ and $y = x + 3$.

By overlaying Figure 5.2.6 and Figure 5.2.7 as in Figure 5.2.8 you can see that Figure 5.2.4 should be given by subtracting $\int_{-1}^2 (x^2 + 1) dx$ from $\int_{-1}^2 (x + 3) dx$.

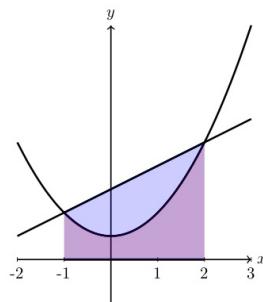


Figure 5.2.8: Graphs of the curves $y = x^2 + 1$ and $y = x + 3$.

5.2.4 Exercises

1. (Finding Antiderivatives) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Finding Antiderivatives (Constants)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Finding Antiderivatives (Polynomials)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Finding Antiderivatives) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Finding Antiderivatives (Power Functions)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Finding Antiderivatives (Roots)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Definite Integrals from Antiderivatives) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. (Definite Integrals from Antiderivatives (Trigonometric Functions)) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
9. (Finding exact displacement) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
10. (Evaluating the definite integral of a rational function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
11. (Evaluating the definite integral of a linear function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
12. (Evaluating the definite integral of a quadratic function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

13. (Simplifying an integrand before integrating) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

14. (Evaluating the definite integral of a trigonometric function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

15. The instantaneous velocity (in meters per minute) of a moving object is given by the function v as pictured in Figure 5.2.9. Assume that on the interval $0 \leq t \leq 4$, $v(t)$ 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.

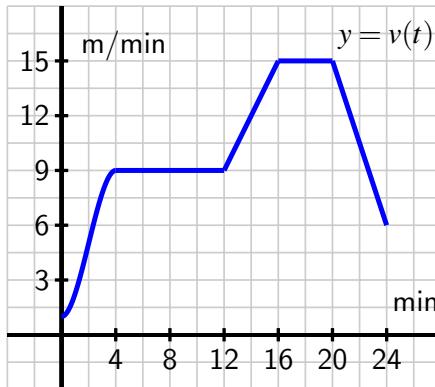


Figure 5.2.9: The velocity function of a moving body.

- Determine the exact distance traveled by the object on the time interval $0 \leq t \leq 4$.
- What is the object's average velocity on $[12, 24]$?
- At what time is the object's acceleration greatest?
- 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?

Answer.

- 20 meters.
- $v_{\text{AVG}[12,24]} = 12.5$ meters per minute.
- The object's maximum acceleration is 3 meters per minute per minute at the instant $t = 2$.
- $c = \frac{60}{24} = 2.5$.

Solution.

- Since velocity is positive, distance traveled is the definite integral of velocity. Thus, on $[0, 4]$, the total distance, D , is

$$D = \int_0^4 v(t) dt = \int_0^4 \left(-\frac{1}{4}t^3 + \frac{3}{2}t^2 + 1 \right) dt.$$

By the Fundamental Theorem of Calculus,

$$\begin{aligned}\int_0^4 \left(-\frac{1}{4}t^3 + \frac{3}{2}t^2 + 1 \right) dt &= \left(-\frac{1}{16}t^4 + \frac{1}{2}t^3 + t \right) \Big|_{t=0}^{t=4} \\ &= -\frac{1}{16}4^4 + \frac{1}{2}4^3 + 4 \\ &= 20,\end{aligned}$$

so the distance traveled by the object on $[0, 4]$ is 20 meters.

- b. To find the average velocity of the object on $[12, 24]$, we first evaluate $\int_{12}^{24} v(t) dt$. To do so, we use the fact that v is piecewise linear and employ standard geometric formulas to determine the value of the integral. The area of the trapezoid formed by $v(t)$ and the vertical line segments at $t = 12$ and $t = 16$ is

$$A_1 = \frac{1}{2}(f(12) + f(16))(4) = 2(9 + 15) = 48.$$

The area of the rectangle formed by $v(t)$ and the vertical line segments at $t = 16$ and $t = 20$ is

$$A_2 = 4 \cdot 15 = 60.$$

Finally, the area of the trapezoid formed by $v(t)$ and the vertical line segments at $t = 20$ and $t = 24$ is

$$A_3 = \frac{1}{2}(f(20) + f(24))(4) = 2(15 + 6) = 42.$$

Thus,

$$\int_{12}^{24} v(t) dt = A_1 + A_2 + A_3 = 48 + 60 + 42 = 150,$$

from which it follows that

$$v_{\text{AVG}[12,24]} = \frac{1}{24 - 12} \int_{12}^{24} v(t) dt = \frac{150}{12} = 12.5$$

meters per minute.

- c. To find where $a(t)$ is greatest, we recall that $a(t) = v'(t)$, so we are seeking the location where the slope of the tangent line to $v(t)$ is maximized on the interval $[0, 24]$. From the graph, there are only two possible intervals where this can occur: $0 < t < 4$ and $12 < t < 16$, as these are both where v is increasing and thus $a(t) = v'(t)$ is positive. On $12 < t < 16$, $a(t) = v'(t) = \frac{3}{2}$ since this is the slope of the line that forms $v(t)$ on this interval. On $0 < t < 4$, $v(t) = -\frac{1}{4}t^3 + \frac{3}{2}t^2 + 1$, so $a(t) = v'(t) = -\frac{3}{4}t^2 + 3t$. We see that $a(t)$ is a quadratic function that opens down, so its maximum occurs at its vertex, which is located where $a'(t) = 0$. We know that $a'(t) = -\frac{3}{2}t + 3$, and $0 = -\frac{3}{2}t + 3$ implies $t = 2$. Moreover, $a(2) = v'(2) = -\frac{3}{4} \cdot 2^2 + 3(2) = 3$. Clearly $a(2) = 3 > \frac{3}{2}$, so the object's maximum acceleration is 3 meters per minute per minute at the instant $t = 2$.
- d. Let $V(t) = v(t) + c$ be the new velocity function that is the original velocity v increased by c at every time. We want to know the value of c for which total distance traveled is 210 meters. For this to be true, we must have that

$$210 = \int_0^{24} V(t) dt = \int_0^{24} (v(t) + c) dt.$$

Using the additive property of the definite integral and our earlier work in (b), we can say that

$$210 = \int_0^{24} v(t) dt + \int_0^{24} c dt = 150 + \int_0^{24} c dt,$$

so

$$\int_0^{24} c dt = 60.$$

Since the function $f(t) = c$ is constant, the integral $\int_0^{24} c dt$ represents the area of a rectangle 24 units wide and c units tall, so $\int_0^{24} c dt = 24c$. Hence,

$$24c = 60$$

$$\text{so } c = \frac{60}{24} = 2.5.$$

16. A function f is given piecewise by the formula

$$f(x) = \begin{cases} -x^2 + 2x + 1, & \text{if } 0 \leq x < 2 \\ -x + 3, & \text{if } 2 \leq x < 3 \\ x^2 - 8x + 15, & \text{if } 3 \leq x \leq 5 \end{cases}.$$

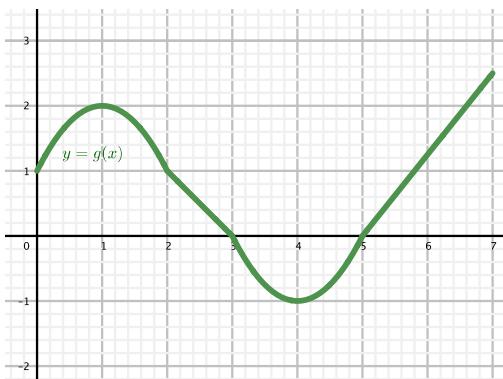
- Determine the exact value of the net signed area enclosed by f and the x -axis on the interval $[2, 5]$.
- Compute the exact average value of f on $[0, 5]$.
- Find a formula for a function g on $5 \leq x \leq 7$ so that if we extend the above definition of f so that $f(x) = g(x)$ if $5 \leq x \leq 7$, it follows that $\int_0^7 f(x) dx = 0$.

Answer.

a. $-\frac{5}{6}$.

b. $f_{\text{AVG}}[0,5] = \frac{1}{2}$.

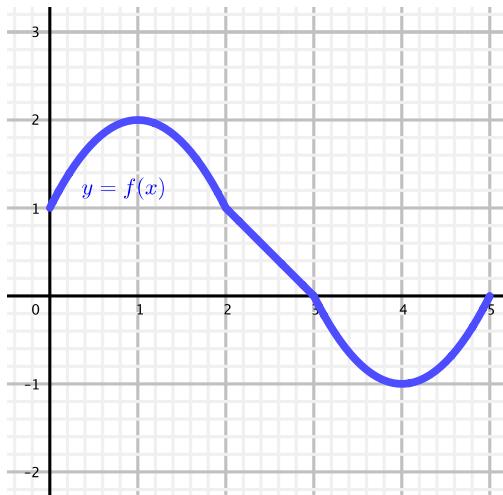
c. $g(x) = f(x)$ for $0 \leq x < 5$ and $g(x) = -\frac{5}{4}(x - 5)$ on $5 \leq x \leq 7$.



Solution.

- a. The net signed area enclosed by f and the x -axis on the interval $[2, 5]$ is $\int_2^5 f(x) dx$. The figure below shows a graph of f on the interval $[0, 5]$. To calculate the definite integral of f on the interval $[2, 5]$, we consider the integrals on the subintervals $[2, 3]$ and $[3, 5]$. Doing so and using the Fundamental Theorem of Calculus,

$$\begin{aligned}
 \int_2^5 f(x) dx &= \int_2^3 f(x) dx + \int_3^5 f(x) dx \\
 &= \int_2^3 -x + 3 dx + \int_3^5 x^2 - 8x + 15 dx \\
 &= \left(-\frac{1}{2}x^2 + 3x \right) \Big|_2^3 + \left(\frac{1}{3}x^3 - 4x^2 + 15x \right) \Big|_3^5 \\
 &= \left(\frac{9}{2} - 4 \right) + \left(\frac{50}{3} - 18 \right) \\
 &= \frac{127}{6} - 22 = -\frac{5}{6}.
 \end{aligned}$$



- b. To find the average value of f on $[0, 5]$ we need to first calculate

$$\int_0^5 f(x) dx.$$

Using our work in (a) and considering the subintervals $[0, 2]$ and $[2, 5]$,

$$\begin{aligned}
 \int_0^5 f(x) dx &= \int_0^2 f(x) dx + \int_2^5 f(x) dx \\
 &= \int_0^2 -x^2 + 2x + 1 dx - \frac{5}{6} \\
 &= \left(-\frac{1}{3}x^3 + x^2 + x \right) \Big|_0^2 - \frac{5}{6} \\
 &= \frac{10}{3} - \frac{5}{6}
 \end{aligned}$$

$$= \frac{5}{2}.$$

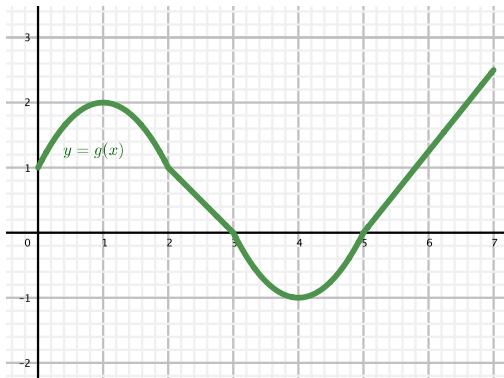
Therefore, the average value of f on $[0, 5]$ is

$$f_{\text{AVG}[0,5]} = \frac{1}{5-0} \int_0^5 f(x) dx = \left(\frac{1}{5}\right) \left(\frac{5}{2}\right) = \frac{1}{2}.$$

- c. Since $\int_0^5 f(x) dx = \frac{5}{2}$, we need a function g that matches f but that is also defined on $[5, 7]$ so that $\int_5^7 g(x) dx = -\frac{5}{2}$. To preserve continuity, we will choose a formula for g that also satisfies $g(5) = 0$. It's natural to pick a linear function whose graph encloses $5/2$ units of area, so one that has a slope of $5/4$. To check, we compute

$$\begin{aligned} \int_5^7 -\frac{5}{4}(x-5) dx &= -\left(\frac{5}{4}\right) \left(\frac{x^2}{2} - 5x\right) \Big|_5^7 \\ &= -\left(\frac{5}{4}\right) \left(\frac{49-25}{2} - 5(7-5)\right) \\ &= -\left(\frac{5}{4}\right) (12-10) \\ &= -\left(\frac{5}{4}\right) (2) = -\frac{5}{2}. \end{aligned}$$

So g defined so that $g(x) = f(x)$ for $0 \leq x < 5$ and $g(x) = -\frac{5}{4}(x-5)$ on $5 \leq x \leq 7$ has the desired property. A graph of g is shown in the figure below.



5.3 Differential Equations

Motivating Questions

- What is a differential equation and what kinds of information can it tell us?
 - How do differential equations arise in the world around us?
 - What do we mean by a solution to a differential equation?
-

In previous chapters, we have seen that a function's derivative tells us the rate at which the function is changing. The Fundamental Theorem of Calculus helped us determine the total change of a function over an interval from the function's rate of change. For instance, an object's velocity tells us the rate of change of that object's position. By integrating the velocity over a time interval, we can determine how much the position changes over that time interval. If we know where the object is at the beginning of that interval, we have enough information to predict where it will be at the end of the interval.

In this chapter, we introduce the concept of *differential equations*. A differential equation is an equation that provides a description of a function's derivative, which means that it tells us the function's rate of change. Using this information, we would like to learn as much as possible about the function itself. Ideally we would like to have an algebraic description of the function. As we'll see, this may be too much to ask in some situations, but we will still be able to make accurate approximations.

5.3.1 What is a differential equation?

A differential equation is an equation that describes the derivative, or derivatives, of a function that is unknown to us. For instance, the equation

$$\frac{dy}{dx} = 2x \sin(x^2)$$

describes the derivative of a function $y(x)$ that is unknown to us.

As many important examples of differential equations involve quantities that change in time, the independent variable in our discussion will frequently be time t . For example, the differential equation

$$\frac{ds}{dt} = 4t + 1$$

describes an object's position ($s(t)$) as a function of time.

If we know the velocity and the starting position of a moving object, we will be able to find its position at any later time.

Because differential equations describe the derivative of a function, they give us information about how that function changes. Our goal will be to use this information to predict the value

of the function in the future; in this way, differential equations provide us with something like a crystal ball.

Example 5.3.1. The position of a moving object is given by the function $s(t)$, where s is measured in feet and t in seconds. We determine that the velocity is $v(t) = 4t + 1$ feet per second.

- How much does the position change over the time interval $[0, 4]$?
- Does this give you enough information to determine $s(4)$, the position at time $t = 4$? If so, what is $s(4)$? If not, what additional information would you need to know to determine $s(4)$?
- Suppose you are told that the object's initial position $s(0) = 7$. Determine $s(2)$, the object's position 2 seconds later.
- If we know that $s(0) = 7$ can we find an equation for $s(t)$?

Hint. What does the quantity $\int_0^4 (4t + 1)dt$ represent in this problem?

Solution.

- We can find the total change in the position of the object by integrating the equation for velocity over the interval $[0, 4]$. That is

$$\int_0^4 (4t + 1)dt = (2t^2 + t)|_0^4 = 36.$$

Therefore, the position changes by 36 feet over the interval $[0, 4]$.

- Simply knowing how much the position has changed on the interval $[0, 4]$ does not tell us where the object is. In order to know where the object is located we must also know where the object started.
- We can find the total change in the position of the object by integrating the equation for velocity over the interval $[0, 2]$. That is

$$\int_0^2 (4t + 1)dt = (2t^2 + t)|_0^2 = 10.$$

Therefore, the position changes by 10 feet over the interval $[0, 2]$. Since we know that $s(0) = 7$ we can say that $s(2) = 7 + 10 = 17$.

- Note that

$$\int (4t + 1)dt = 2t^2 + t + C.$$

Since $s(0) = 7$ it is reasonable to use $s(0)$ to "find" the constant C . That is, we can say that $7 = 2(0)^2 + 0 + C$ which means $C = 7$. Therefore, the position of the object at anytime is given by the equation $s(t) = 2t^2 + t + 7$.

5.3.2 Solving a differential equation

A differential equation describes the derivative, or derivatives, of a function that is unknown to us. By a *solution* to a differential equation, we mean simply a function that satisfies this description.

For instance, the first differential equation we looked at is

$$\frac{ds}{dt} = 4t + 1,$$

which describes an unknown function $s(t)$. We may check that $s(t) = 2t^2 + t$ is a solution because it satisfies this description. Notice that $s(t) = 2t^2 + t + 4$ is also a solution.

If we have a candidate for a solution, it is straightforward to check whether it is a solution or not. Before we demonstrate, however, let's consider the same issue in a simpler context. Suppose we are given the equation $2\Box^2 - 2\Box = 2\Box + 6$ and asked whether $\Box = 3$ is a solution. To answer this question, we could rewrite the variable \Box in the equation with the symbol \Box :

$$2\Box^2 - 2\Box = 2\Box + 6.$$

To determine whether $\Box = 3$ is a solution, we can investigate the value of each side of the equation separately when the value 3 is placed in \Box and see if indeed the two resulting values are equal. Doing so, we observe that

$$2\Box^2 - 2\Box = 2 \cdot 3^2 - 2 \cdot 3 = 12,$$

and

$$2\Box + 6 = 2 \cdot 3 + 6 = 12.$$

Therefore, $\Box = 3$ is indeed a solution.

We will do the same thing with differential equations. Consider the differential equation

$$\begin{aligned} \frac{dv}{dt} &= 1.5 - 0.5v, \text{ or} \\ \frac{d\Box}{dt} &= 1.5 - 0.5\Box. \end{aligned}$$

Let's ask whether $v(t) = 3 - 2e^{-0.5t}$ is a solution¹. Using this formula for v , observe first that

$$\frac{dv}{dt} = \frac{d\Box}{dt} = \frac{d}{dt}[3 - 2e^{-0.5t}] = -2e^{-0.5t} \cdot (-0.5) = e^{-0.5t}$$

and

$$1.5 - 0.5v = 1.5 - 0.5\Box = 1.5 - 0.5(3 - 2e^{-0.5t}) = 1.5 - 1.5 + e^{-0.5t} = e^{-0.5t}.$$

Since $\frac{dv}{dt}$ and $1.5 - 0.5v$ agree for all values of t when $v = 3 - 2e^{-0.5t}$, we have indeed found a solution to the differential equation.

¹At this time, don't worry about why we chose this function; we will learn techniques for finding solutions to differential equations soon enough.

Example 5.3.2. Consider the differential equation

$$\frac{dv}{dt} = 1.5 - 0.5v.$$

Which of the following functions are solutions of this differential equation?

- a. $v(t) = 1.5t - 0.25t^2$.
- b. $v(t) = 3 + 2e^{-0.5t}$.
- c. $v(t) = 3$.
- d. $v(t) = 3 + Ce^{-0.5t}$ where C is any constant.

Hint.

- a. hints for each of the prompts above.

Answer.

- a. $v(t) = 1.5t - 0.25t^2$ is not a solution to the given DE.
- b. $v(t) = 3 + 2e^{-0.5t}$ is a solution to the given DE.
- c. $v(t) = 3$ is a solution to the given DE.
- d. $v(t) = 3 + Ce^{-0.5t}$ is a solution to the given DE for any choice of C .

Solution.

- a. Since $v(t) = 1.5t - 0.25t^2$, $\frac{dv}{dt} = 1.5 - 0.5t$. We observe that for this choice of v , $1.5 - 0.5v = 1.5 - 0.5(1.5t - 0.25t^2) = 1.5 - 0.75t + 0.125t^2 \neq 1.5 - 0.5t = \frac{dv}{dt}$, and thus $v(t) = 1.5t - 0.25t^2$ is not a solution to the given differential equation.
- b. Since $v(t) = 3 + 2e^{-0.5t}$, $\frac{dv}{dt} = 2e^{-0.5t}(-0.5) = -e^{-0.5t}$. Hence, for this choice of v , $1.5 - 0.5v = 1.5 - 0.5(3 + 2e^{-0.5t}) = 1.5 - 1.5 - e^{-0.5t} = -e^{-0.5t} = \frac{dv}{dt}$. Therefore $v(t) = 3 + 2e^{-0.5t}$ is a solution to the given differential equation.
- c. Since $v(t) = 3$, $\frac{dv}{dt} = 0$. Hence, for this choice of v , $1.5 - 0.5v = 1.5 - 0.5(3) = 1.5 - 1.5 = 0 = \frac{dv}{dt}$. Therefore $v(t) = 3$ is a solution to the given differential equation.
- d. Since $v(t) = 3 + Ce^{-0.5t}$ (where C is a constant), $\frac{dv}{dt} = Ce^{-0.5t}(-0.5) = -0.5Ce^{-0.5t}$. Hence, for this choice of v , $1.5 - 0.5v = 1.5 - 0.5(3 + Ce^{-0.5t}) = 1.5 - 1.5 - 0.5Ce^{-0.5t} = -0.5Ce^{-0.5t} = \frac{dv}{dt}$. Therefore $v(t) = 3 + Ce^{-0.5t}$ is a solution to the given differential equation for any choice of C .

This example shows us something interesting. Notice that the differential equation has infinitely many solutions, which are parametrized by the constant C in $v(t) = 3 + Ce^{-0.5t}$. In Figure 5.3.3, we see the graphs of these solutions for a few values of C , as labeled.

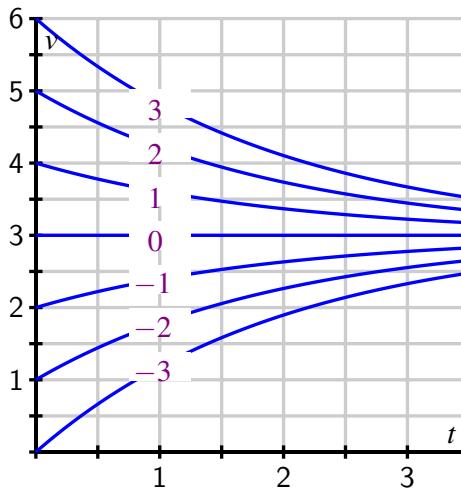


Figure 5.3.3: The family of solutions to the differential equation $\frac{dv}{dt} = 1.5 - 0.5v$.

Notice that the value of C is connected to the initial value of the velocity $v(0)$, since $v(0) = 3 + C$. In other words, while the differential equation describes how the velocity changes as a function of the velocity itself, this is not enough information to determine the velocity uniquely: we also need to know the initial velocity. For this reason, differential equations will typically have infinitely many solutions, one corresponding to each initial value. We have seen this phenomenon before: given the velocity of a moving object $v(t)$, we cannot uniquely determine the object's position function unless we also know its initial position.

If we are given a differential equation and an initial value for the unknown function, we say that we have an *initial value problem*. For instance,

$$\frac{dv}{dt} = 1.5 - 0.5v, \quad v(0) = 0.5$$

is an initial value problem. In this problem, we know the value of v at one time and we know how v is changing. Consequently, there should be exactly one function v that satisfies the initial value problem.

This demonstrates the following important general property of initial value problems.

Initial value problems that are “well behaved” have exactly one solution, which exists in some interval around the initial point.

We won’t worry about what “well behaved” means —it is a technical condition that will be satisfied by all the differential equations we consider.

5.3.3 Exercises

1. (General Solution to a Differential Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

2. (Developing a Differential Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Developing a Differential Equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Displacement) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Initial Value Problems) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Initial Velocity) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. (Finding the Position Function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
8. (Finding the Position Function Given Acceleration) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
9. (Finding Discplacement) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
10. (Finding Discplacement) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
11. (Finding Discplacement from a Graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
12. (Finding Discplacement from a Graph) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

5.4 The Second Fundamental Theorem of Calculus

Motivating Questions

- How does the integral function $A(x) = \int_1^x f(t) dt$ define an antiderivative of f ?
 - What is the statement of the Second Fundamental Theorem of Calculus?
 - How do the First and Second Fundamental Theorems of Calculus enable us to formally see how differentiation and integration are almost inverse processes?
-

In Section 4.4, we learned the Fundamental Theorem of Calculus (FTC), which from here forward will be referred to as the *First* Fundamental Theorem of Calculus, as in this section we develop a corresponding result that follows it. Recall that the First FTC tells us that if f is a continuous function on $[a, b]$ and F is any antiderivative of f (that is, $F' = f$), then

$$\int_a^b f(x) dx = F(b) - F(a).$$

We have used this result in two settings:

- 1 If we have a graph of f and we can compute the exact area bounded by f on an interval $[a, b]$, we can compute the change in an antiderivative F over the interval.
- 2 If we can find an algebraic formula for an antiderivative of f , we can evaluate the integral to find the net signed area bounded by the function on the interval.

For the former, see Example 5.1.1 or Example 5.1.3. For the latter, we can easily evaluate exactly integrals such as

$$\int_1^4 x^2 dx,$$

since we know that the function $F(x) = \frac{1}{3}x^3$ is an antiderivative of $f(x) = x^2$. Thus,

$$\begin{aligned} \int_1^4 x^2 dx &= \frac{1}{3}x^3 \Big|_1^4 \\ &= \frac{1}{3}(4)^3 - \frac{1}{3}(1)^3 \\ &= 21. \end{aligned}$$

Thus, the First FTC can be used in two ways. First, to find the difference $F(b) - F(a)$ for an antiderivative F of the integrand f , even if we may not have a formula for F itself. To do this, we must know the value of the integral $\int_a^b f(x) dx$ exactly, perhaps through known geometric formulas for area. In addition, the First FTC provides a way to find the exact value of a

definite integral, and hence a certain net signed area exactly, by finding an antiderivative of the integrand and evaluating its total change over the interval. In this case, we need to know a formula for the antiderivative F . Both of these perspectives are reflected in Figure 5.4.1.

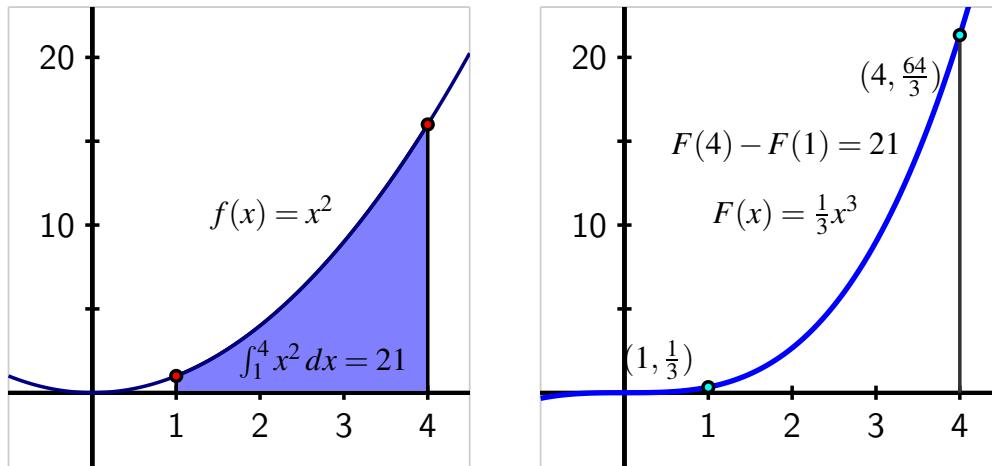


Figure 5.4.1: At left, the graph of $f(x) = x^2$ on the interval $[1, 4]$ and the area it bounds. At right, the antiderivative function $F(x) = \frac{1}{3}x^3$, whose total change on $[1, 4]$ is the value of the definite integral at left.

The value of a definite integral may have additional meaning depending on context: as the change in position when the integrand is a velocity function, the total amount of pollutant leaked from a tank when the integrand is the rate at which pollution is leaking, or other total changes if the integrand is a rate function. Also, the value of the definite integral is connected to the average value of a continuous function on a given interval: $f_{\text{AVG}[a,b]} = \frac{1}{b-a} \int_a^b f(x) dx$.

In the last part of Section 5.1, we studied integral functions of the form $A(x) = \int_c^x f(t) dt$. Figure 5.1.8 is a particularly important image to keep in mind as we work with integral functions, and the corresponding java applet at gvsu.edu/s/cz can help us understand the function A . In what follows, we use the First FTC to gain additional understanding of the function $A(x) = \int_c^x f(t) dt$, where the integrand f is given (either through a graph or a formula), and c is a constant.

Example 5.4.2. Consider the function A defined by the rule

$$A(x) = \int_1^x f(t) dt,$$

where $f(t) = 4 - 2t$.

- Compute $A(1)$ and $A(2)$ exactly.
- Use the First Fundamental Theorem of Calculus to find a formula for $A(x)$ that does not involve integrals. That is, use the first FTC to evaluate $\int_1^x (4 - 2t) dt$.

- c. Observe that f is a linear function; what kind of function is A ?
- d. Using the formula you found in (b) that does not involve integrals, compute $A'(x)$.
- e. 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 ?

5.4.1 The Second Fundamental Theorem of Calculus

The result of Example 5.4.2 is not particular to the function $f(t) = 4 - 2t$, nor to the choice of “1” as the lower bound in the integral that defines the function A . For instance, if we let $f(t) = \cos(t) - t$ and set $A(x) = \int_2^x f(t) dt$, we can determine a formula for A by the First FTC. Specifically,

$$\begin{aligned} A(x) &= \int_2^x (\cos(t) - t) dt \\ &= \sin(t) - \frac{1}{2}t^2 \Big|_2^x \\ &= \sin(x) - \frac{1}{2}x^2 - (\sin(2) - 2). \end{aligned}$$

Differentiating $A(x)$, since $(\sin(2) - 2)$ is constant, it follows that

$$A'(x) = \cos(x) - x,$$

and thus we see that $A'(x) = f(x)$, so A is an antiderivative of f . And since $A(2) = \int_2^2 f(t) dt = 0$, A is the only antiderivative of f for which $A(2) = 0$.

In general, if f is any continuous function, and we define the function A by the rule

$$A(x) = \int_c^x f(t) dt,$$

where c 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

$$\begin{aligned} A'(x) &= \lim_{h \rightarrow 0} \frac{A(x+h) - A(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\int_c^{x+h} f(t) dt - \int_c^x f(t) dt}{h} \\ &= \lim_{h \rightarrow 0} \frac{\int_x^{x+h} f(t) dt}{h}, \end{aligned} \tag{5.4.1}$$

where Equation (5.4.1) follows from the fact that $\int_c^x f(t) dt + \int_x^{x+h} f(t) dt = \int_c^{x+h} f(t) dt$. 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 (5.4.1), it follows that

$$A'(x) = \lim_{h \rightarrow 0} \frac{\int_x^{x+h} f(t) dt}{h} = \lim_{h \rightarrow 0} \frac{f(x) \cdot h}{h} = f(x).$$

Hence, A is indeed an antiderivative of f . In addition, $A(c) = \int_c^c f(t) dt = 0$. The preceding argument demonstrates the truth of the Second Fundamental Theorem of Calculus, which we state as follows.

The Second Fundamental Theorem of Calculus

If f is a continuous function and c is any constant, then f has a unique antiderivative A that satisfies $A(c) = 0$, and that antiderivative is given by the rule $A(x) = \int_c^x f(t) dt$.

Example 5.4.3. Suppose that f is the function given in Figure 5.4.4 and that f is a piecewise function whose parts are either portions of lines or portions of circles, as pictured.

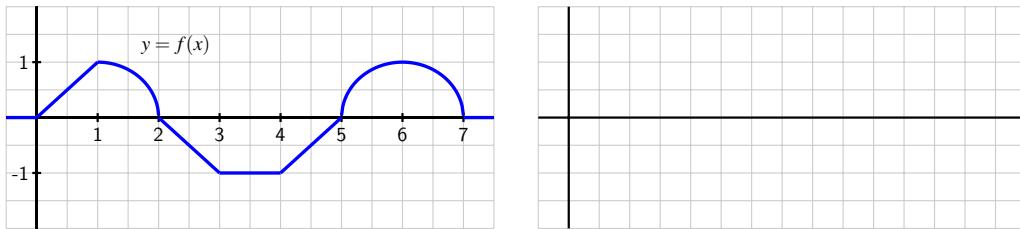


Figure 5.4.4: At left, the graph of $y = f(x)$. At right, axes for sketching $y = A(x)$.

In addition, let A be the function defined by the rule $A(x) = \int_2^x f(t) dt$.

- What does the Second FTC tell us about the relationship between A and f ?
- Compute $A(1)$ and $A(3)$ exactly.
- Sketch a precise graph of $y = A(x)$ on the axes at right that accurately reflects where A is increasing and decreasing, where A is concave up and concave down, and the exact values of A at $x = 0, 1, \dots, 7$.
- How is A similar to, but different from, the function F that you found in Example 5.1.3?
- With as little additional work as possible, sketch precise graphs of the functions $B(x) = \int_3^x f(t) dt$ and $C(x) = \int_1^x f(t) dt$. Justify your results with at least one sentence of explanation.

Hint.

- If you don't recall it, review the statement of the Second FTC above.

- b. Note that $A(1) = \int_2^1 f(t) dt$.
- c. Don't miss our key conclusion from (a).
- d. Compare the values of $A(1)$ and $F(1)$.
- e. What does the Second FTC tell us about the relationship between B and f ?

Answer.

- a. $A'(x) = f(x)$.
- b. $A(1) = -\frac{\pi}{4}$.
- c. A is increasing wherever f is positive; A is CCU wherever f is increasing. $A(2) = 0$, $A(3) = -0.5$, $A(4) = -1.5$, $A(5) = -2$, $A(6) = -2 + \frac{\pi}{4}$, and $A(7) = -2 + \frac{\pi}{2}$.
- d. F and A differ by the constant $\frac{\pi}{4} - \frac{1}{2}$.
- e. B and C have the same shape as A and F , and differ from A by a constant. Observe that $B(3) = 0$ and $C(1) = 0$.

Solution.

- a. By the Second FTC, $A'(x) = f(x)$.
- b. Since $A(1) = \int_2^1 f(t) dt = -\int_1^2 f(t) dt$, it follows $A(1) = -\frac{\pi}{4}$.
- c. Note that A is increasing wherever f is positive, and A is CCU wherever f is increasing. Similar conclusions follow for A being decreasing and/or concave down. Moreover, $A(2) = 0$, $A(3) = -0.5$, $A(4) = -1.5$, $A(5) = -2$, $A(6) = -2 + \frac{\pi}{4}$, and $A(7) = -2 + \frac{\pi}{2}$.
- d. In our current example, A is an antiderivative of f that satisfies $A(0) = -\frac{1}{2} - \frac{\pi}{4}$. Our earlier work with F showed that F is an antiderivative of F that satisfied $F(0) = -1$. Since F and A are both antiderivatives of f , they differ by a constant, and that constant is $-1 - (-\frac{1}{2} - \frac{\pi}{4}) = \frac{\pi}{4} - \frac{1}{2}$.
- e. The Second FTC tells us that $B' = f$ and $C' = f$. Thus, B and C are each antiderivatives of f , have the same shape as A and F , and each differ from A by just a constant. Observing that $B(3) = 0$ and $C(1) = 0$ enables us to easily sketch these shifted versions of A .

5.4.2 Understanding Integral Functions

The Second FTC provides us with a way 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 $G'(x) = g(x)$, or

$$\frac{d}{dx} \left[\int_c^x g(t) dt \right] = g(x). \quad (5.4.2)$$

This result is useful for understanding the graph of G .

Example 5.4.5. Investigate the behavior of the integral function

$$E(x) = \int_0^x e^{-t^2} dt.$$

Solution. E is closely related to the well known *error function*¹ in probability and statistics. It turns out that the function e^{-t^2} does not have an elementary antiderivative.

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 . By applying the rule in Equation (5.4.2) 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 , and we know that $E(0) = 0$. This information is precisely the type we were given in Example 3.1.10, where we were given information about the derivative of a function, but lacked a formula for the function itself.

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, 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, as $x \rightarrow \infty$, $E'(x) = e^{-x^2} \rightarrow 0$, so the slope of the function E tends to zero as $x \rightarrow \infty$ (and similarly as $x \rightarrow -\infty$). Indeed, it turns out 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, which aligns with the fact that E has horizontal asymptotes. Putting all of this information together (and using the symmetry of $f(t) = e^{-t^2}$), we see the results shown in Figure 5.4.6.

¹The error function is defined by the rule $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ and has the key property that $0 \leq \text{erf}(x) < 1$ for all $x \geq 0$ and moreover that $\lim_{x \rightarrow \infty} \text{erf}(x) = 1$.

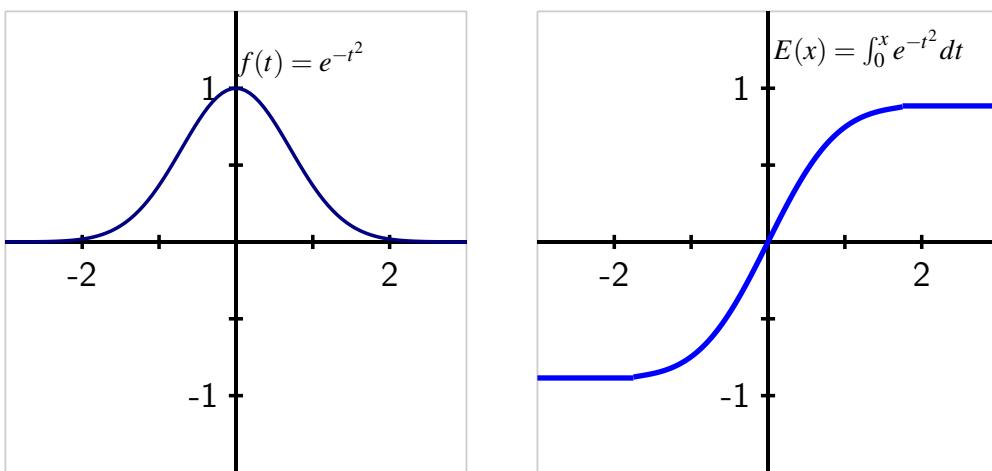


Figure 5.4.6: At left, the graph of $f(t) = e^{-t^2}$. At right, the integral function $E(x) = \int_0^x e^{-t^2} dt$, which is the unique antiderivative of f that satisfies $E(0) = 0$.

Because E is the antiderivative of $f(t) = e^{-t^2}$ that satisfies $E(0) = 0$, 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 as x increases, since there is less and less additional accumulated area bounded by $f(t) = e^{-t^2}$ as x increases.

Example 5.4.7. Suppose that $f(t) = \frac{t}{1+t^2}$ and $F(x) = \int_0^x f(t) dt$.

- On the axes at left in Figure 5.4.8, plot a graph of $f(t) = \frac{t}{1+t^2}$ on the interval $-10 \leq t \leq 10$. Clearly label the vertical axes with appropriate scale.
- What is the key relationship between F and f , according to the Second FTC?
- Use the first derivative test to determine the intervals on which F is increasing and decreasing.
- 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 $f'(t) = \frac{1-t^2}{(1+t^2)^2}$.
- Using technology appropriately, estimate the values of $F(5)$ and $F(10)$ through appropriate Riemann sums.
- Sketch an accurate graph of $y = F(x)$ on the righthand axes provided, and clearly label the vertical axes with appropriate scale.

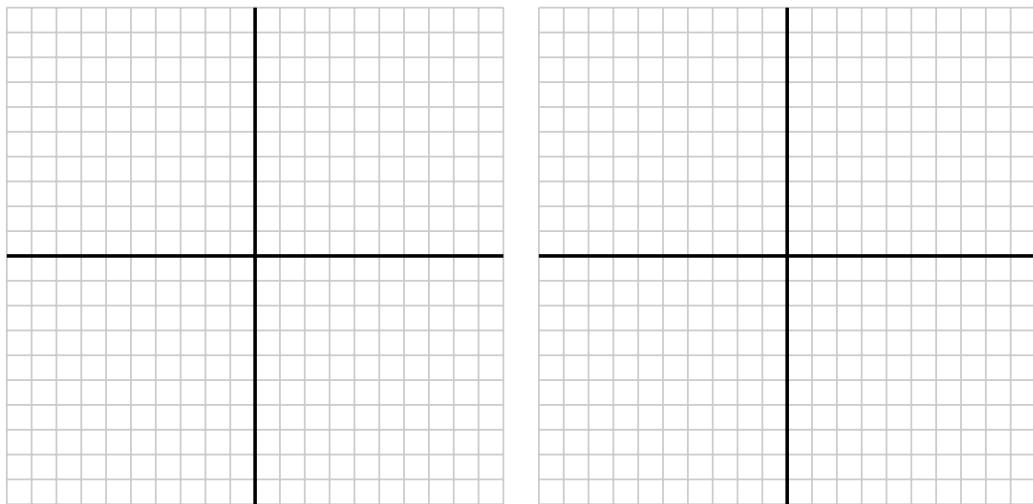


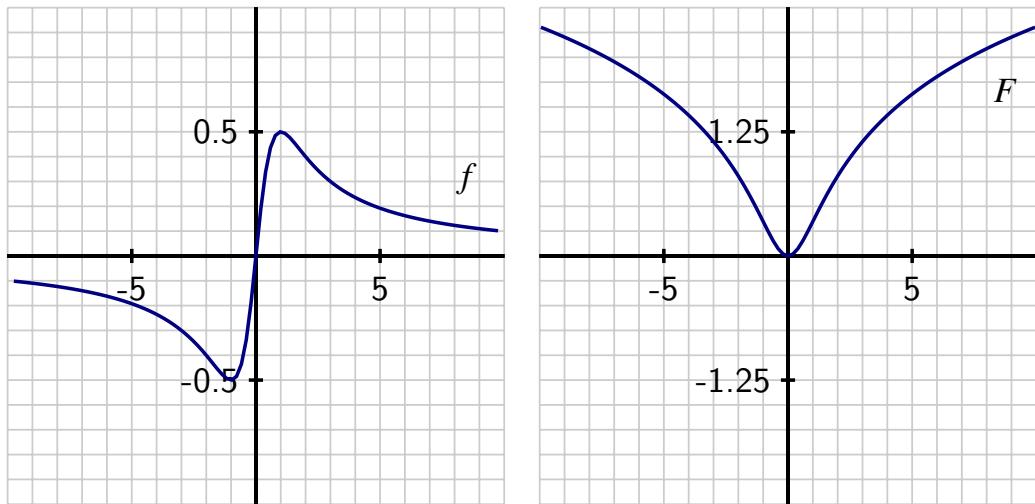
Figure 5.4.8: Axes for plotting f and F .

Hint.

- a. Use computing technology appropriately to generate the desired graph.
- b. Again, recall the statement of the Second FTC.
- c. Where is F' positive? F' negative?
- d. Note that $F'' = f'$.
- e. Remember that $F(5) = \int_0^5 \frac{t}{1+t^2} dt$.
- f. Don't forget that $F(0) = 0$.

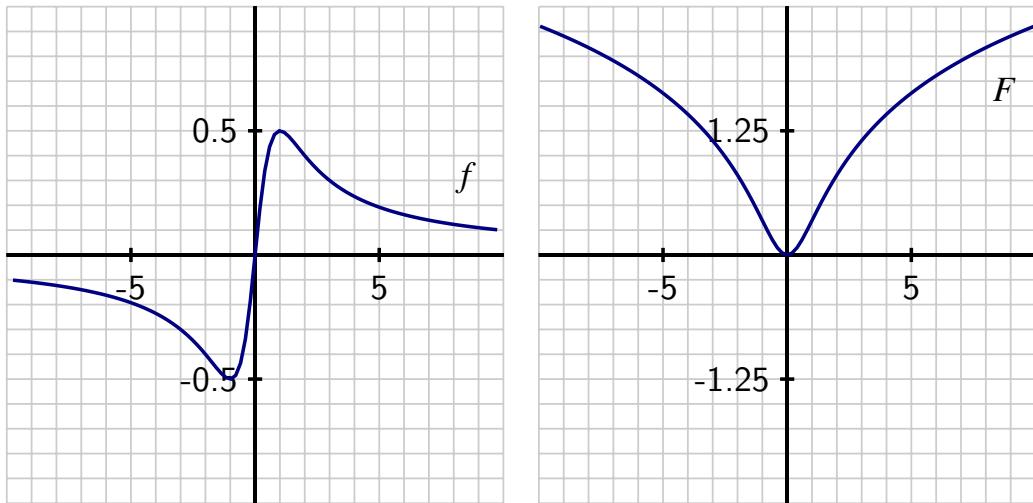
Answer.

- a. See the plot at below left.
- b. $F' = f$.
- c. F is increasing for all $x > 0$; F is decreasing for $x < 0$
- d. F is CCU on $-1 < x < 1$ and CCD for $x < -1$ and $x > 1$.
- e. $F(5) \approx 2.35973$; $F(10) \approx 2.35973$.
- f. See the graph at below right.



Solution.

- a. See the plot at below left.
- b. $F' = f$, by the Second FTC.
- c. F is increasing wherever $F' = f$ is positive, so for all $x > 0$. Similarly, F is decreasing for $x < 0$
- d. F is CCU wherever $F' = f$ is increasing or wherever $F'' = f'$ is positive. It is straightforward to show that f'' is positive for $-1 < x < 1$ and negative otherwise, thus F is CCU on $-1 < x < 1$ and CCD for $x < -1$ and $x > 1$.
- e. $F(5) = \int_0^5 \frac{t}{1+t^2} dt \approx 2.35973$, using a midpoint Riemann sum with 10 subintervals. Similarly, $F(10) = \int_0^{10} \frac{t}{1+t^2} dt \approx 2.35973$.
- f. Recalling that $F(0) = 0$ and using the values and information we've found in (b)-(e), we arrive at the graph at below right.



5.4.3 Differentiating an Integral Function

We have seen that the Second FTC enables us to construct an antiderivative F for any continuous function f as the integral function $F(x) = \int_c^x f(t) dt$. If we have a function of the form $F(x) = \int_c^x f(t) dt$, then we know that $F'(x) = \frac{d}{dx} \left[\int_c^x f(t) dt \right] = f(x)$. 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

$$F(x) = \int_{\pi}^x \sin(t^2) dt,$$

then by the Second FTC, we know immediately that

$$F'(x) = \sin(x^2).$$

In general, we know by the Second FTC that

$$\frac{d}{dx} \left[\int_a^x f(t) dt \right] = f(x).$$

This equation says that “the derivative of the integral function whose integrand is f , is f .“ 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“ each other.

What happens if we differentiate a function $f(t)$ and then integrate the result from $t = a$ to $t = x$? That is, what can we say about the quantity

$$\int_a^x \frac{d}{dt} [f(t)] dt?$$

We note that $f(t)$ is an antiderivative of $\frac{d}{dt}[f(t)]$ and apply the First FTC. We see that

$$\begin{aligned}\int_a^x \frac{d}{dt}[f(t)] dt &= f(t) \Big|_a^x \\ &= f(x) - f(a).\end{aligned}$$

Thus, we see that if we first differentiate f and then integrate the result from a to x , we return to the function f , minus the constant value $f(a)$. So the two processes almost undo each other, 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. This should not be surprising: integrating involves antiderivatives, which reverses the process of differentiating. On the other hand, we see that there is some subtlety involved, because integrating the derivative of a function does not quite produce the function itself. This is because every function has an entire family of antiderivatives, and any two of those antiderivatives differ only by a constant.

Example 5.4.9. Evaluate each of the following derivatives and definite integrals. Clearly cite whether you use the First or Second FTC in so doing.

a. $\frac{d}{dx} \left[\int_4^x e^{t^2} dt \right]$

d. $\int_3^x \frac{d}{dt} [\ln(1+t^2)] dt$

b. $\int_{-2}^x \frac{d}{dt} \left[\frac{t^4}{1+t^4} \right] dt$

c. $\frac{d}{dx} \left[\int_x^1 \cos(t^3) dt \right]$

e. $\frac{d}{dx} \left[\int_4^{x^3} \sin(t^2) dt \right].$

Hint.

a. Second FTC.

b. First FTC.

c. $\int_x^1 g(t) dt = - \int_1^x g(t) dt.$

d. First FTC.

e. Let $F(x) = \int_4^x \sin(t^2) dt$ and observe that this problem is asking you to evaluate $\frac{d}{dx} [F(x^3)]$.

Answer.

a. $\frac{d}{dx} \left[\int_4^x e^{t^2} dt \right] = e^{x^2}.$

b. $\int_{-2}^x \frac{d}{dt} \left[\frac{t^4}{1+t^4} \right] dt = \frac{x^4}{1+x^4} - \frac{8}{9}.$

c. $\frac{d}{dx} \left[\int_x^1 \cos(t^3) dt \right] = -\cos(x^3).$

d. $\int_3^x \frac{d}{dt} [\ln(1+t^2)] dt = \ln(1+x^2) - \ln(4).$

e. $\frac{d}{dx} \left[\int_4^{x^3} \sin(t^2) dt \right] = \sin(x^6) \cdot 3x^2.$

Solution.

- By the Second FTC, $\frac{d}{dx} \left[\int_4^x e^{t^2} dt \right] = e^{x^2}.$
- By the First FTC, $\int_{-2}^x \frac{d}{dt} \left[\frac{t^4}{1+t^4} \right] dt = \left[\frac{t^4}{1+t^4} \right] \Big|_{-2}^x = \frac{x^4}{1+x^4} - \frac{8}{9}.$
- Since $\int_x^1 g(t) dt = - \int_1^x g(t) dt$, it follows by this fact and the Second FTC that $\frac{d}{dx} \left[\int_x^1 \cos(t^3) dt \right] = - \frac{d}{dx} \left[\int_1^x \cos(t^3) dt \right] = -\cos(x^3).$
- By the First FTC, $\int_3^x \frac{d}{dt} [\ln(1+t^2)] dt = \ln(1+t^2) \Big|_3^x = \ln(1+x^2) - \ln(4).$
- Letting $F(x) = \int_4^x \sin(t^2) dt$ it follows that we need to compute $\frac{d}{dx} [F(x^3)].$ By the Chain Rule, $\frac{d}{dx} [F(x^3)] = F'(x^3) \cdot 3x^2.$ By the Second FTC, we know that $F'(x) = \sin(x^2)$, and thus $\frac{d}{dx} \left[\int_4^{x^3} \sin(t^2) dt \right] = \frac{d}{dx} [F(x^3)] = \sin(x^6) \cdot 3x^2.$

5.4.4 Summary

- For a continuous function f , the integral function $A(x) = \int_1^x f(t) dt$ defines an antiderivative of f .
- The Second Fundamental Theorem of Calculus is the formal, more general statement of the preceding fact: if f is a continuous function and c is any constant, then $A(x) = \int_c^x f(t) dt$ is the unique antiderivative of f that satisfies $A(c) = 0$.
- Together, the First and Second FTC enable us to formally see how differentiation and integration are almost inverse processes through the observations that

$$\int_c^x \frac{d}{dt} [f(t)] dt = f(x) - f(c)$$

and

$$\frac{d}{dx} \left[\int_c^x f(t) dt \right] = f(x).$$

5.4.5 Exercises

- (A definite integral starting at 3) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Variable in the lower limit) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Approximating a function with derivative $e^{-x^2/5}$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

4. Let g be the function pictured at left in Figure 5.4.10, and let F be defined by $F(x) = \int_2^x g(t) dt$. Assume that the shaded areas have values $A_1 = 4.29$, $A_2 = 12.75$, $A_3 = 0.36$, and $A_4 = 1.79$. Assume further that the portion of A_2 that lies between $x = 0.5$ and $x = 2$ is 6.06. Sketch a carefully labeled graph of F on the axes provided, and include a written analysis of how you know where F is zero, increasing, decreasing, CCU, and CCD.

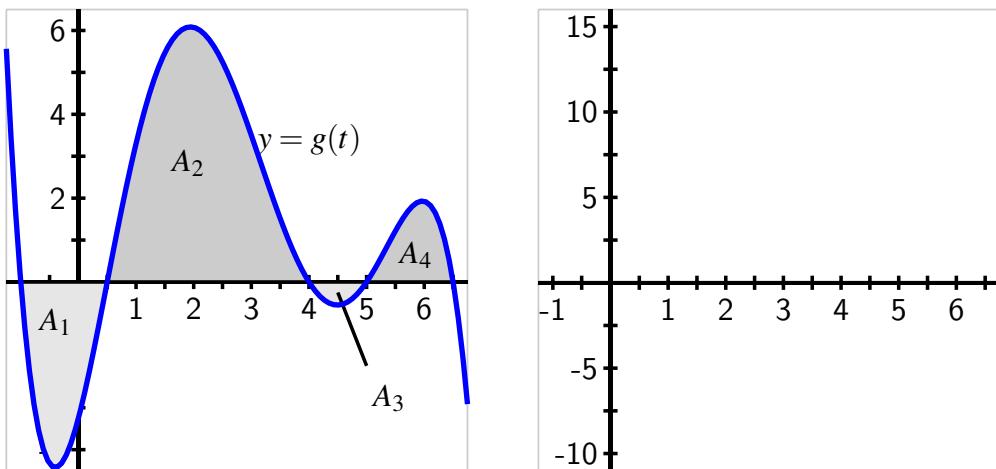
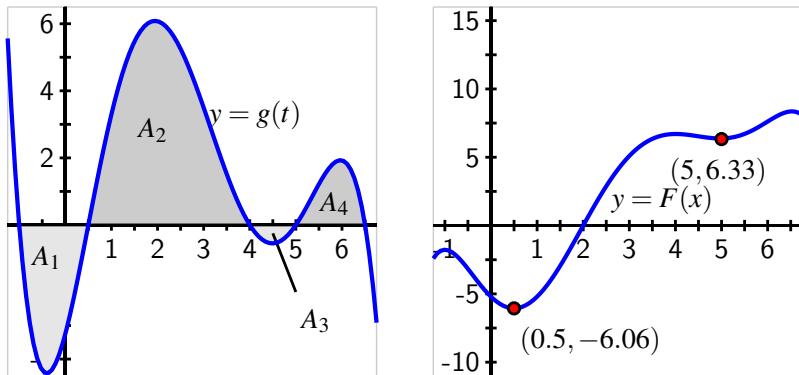


Figure 5.4.10: At left, the graph of g . At right, axes for plotting F .

Answer. F is increasing on $x < -1$, $0.5 < x < 4$, and $5 < x < 6.5$; decreasing on $-1 < x < 0.5$ and $4 < x < 5$; concave up on approximately $-0.4 < x < 2$ and $4.5 < x < 6$; concave down on approximately $2 < x < 4.5$ and $x > 6$; $F(2) = 0$; $F(0.5) = -6.06$; $F(-1) = -1.77$; $F(4) = 6.69$; $F(5) = 6.33$; $F(6.5) = 8.12$.



Solution. By the Second Fundamental Theorem of Calculus, we know that $F'(x) = g(x)$. Thus, from the given graph of g , it follows that F is increasing wherever g is positive (on the intervals $x < -1$, $0.5 < x < 4$, and $5 < x < 6.5$) and decreasing wherever g is negative (on $-1 < x < 0.5$ and $4 < x < 5$). Furthermore, F is concave up wherever g is increasing (on

approximately $-0.4 < x < 2$ and $4.5 < x < 6$) and concave down wherever g is decreasing (on approximately $2 < x < 4.5$ and $x > 6$).

In addition, from the given data about definite integrals and properties of the definite integral, we can determine the following function values:

$$F(2) = \int_2^2 g(t) dt = 0$$

$$F(0.5) = \int_2^{0.5} g(t) dt = -6.06$$

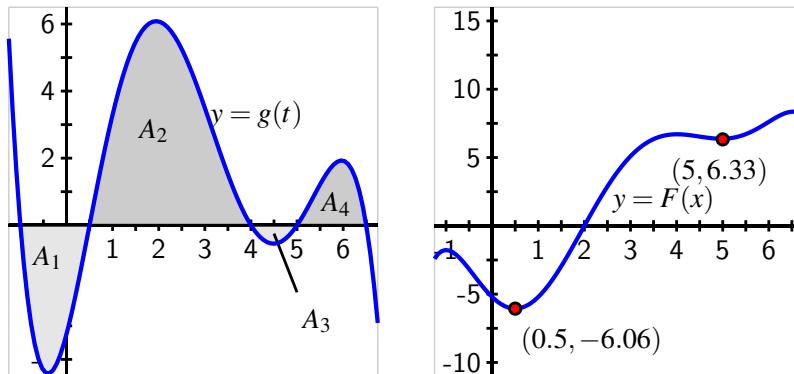
$$F(4) = \int_2^4 g(t) dt = \int_{0.5}^4 g(t) dt - \int_{0.5}^2 g(t) dt = 12.75 - 6.06 = 6.69$$

$$F(5) = F(4) + \int_4^5 g(t) dt = 6.69 - 0.36 = 6.33$$

$$F(6.5) = F(5) + \int_4^{6.5} g(t) dt = 6.33 + 1.79 = 8.12$$

$$F(-1) = \int_2^{-1} g(t) dt = - \int_{-1}^2 g(t) dt = - \left(\int_{-1}^{0.5} g(t) dt + \int_{0.5}^2 g(t) dt \right) = -(-4.29 + 6.06) = -1.77$$

In the figure below, we put all of this information to work and sketch an approximate graph of $y = F(x)$.



5. The tide removes sand from 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 $0 \leq t \leq 6$. 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 $0 \leq t \leq 6$? 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 $0 \leq t \leq 6$, at what time t is the amount of sand on the beach least? What is this minimum value? Explain and justify your answers fully.

Answer. Exercise Answer

Solution.

- a. Since the tide removes sand at a rate of $R(t) = 2 + 5 \sin\left(\frac{4\pi t}{25}\right)$ cubic yards of sand per hour, on the time interval $[0, 6]$, it follows that since the units on $R(t) \cdot \Delta t$ are “cubic yards”, the total sand removed on this time interval is

$$\int_0^6 \left[2 + 5 \sin\left(\frac{4\pi t}{25}\right) \right] dt.$$

- b. Because sand is being both removed and added, we can determine the total amount of sand on the beach at a given time t by integrating the difference between the the rates at which sand is being added and removed. In particular, since the total change is the integral of the rate of change, we have that the total amount of sand on the beach at time x is given by

$$Y(x) = \int_0^x [S(t) - R(t)] dt = \int_0^x \left[\frac{15t}{1+3t} - \left(2 + 5 \sin\left(\frac{4\pi t}{25}\right) \right) \right] dt.$$

- c. We know that $Y(x)$ measures the total amount of sand on the beach at time x . Thus, the instantaneous rate at which the total number of cubic yards of sand on the beach is changing at time $t = 4$ is given by $Y'(4)$. We first compute $Y'(x)$. By the Second Fundamental Theorem of Calculus, we have

$$Y'(x) = S(x) - R(x),$$

and thus $Y'(4) = S(4) - R(4) \approx -1.90875$ cubic yards per hour.

- d. To find where the total sand function $Y(x)$ has its minimum value on $[0, 6]$, we look at the behavior of $Y'(x) = S(x) - R(x)$. Plotting $Y'(x)$ on the interval, we see that $Y'(x) = 0$ at $x \approx 5.118$, and for $0 \leq x < 5.118$, $Y'(x) < 0$, and for $5.118 < x \leq 6$, $Y'(x) > 0$. Thus, Y changes from decreasing to increasing at $x \approx 5.118$ and has an absolute minimum there on the interval $[0, 6]$. Furthermore, $Y(5.118) \approx 2492.368$.
- 6. 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	6000	7000	8000	9000	10,000
c (ft/min)	925	875	830	780	730	685	635	585	535	490	440

Table 5.4.11: Data for the climbing aircraft.

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.

- 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 .
- 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.
- 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.
- 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.
- Estimate the values of $M(6000)$ and $M(10000)$ as accurately as you can. Include units on your results.

Answer. Exercise Answer

Solution.

- Since the given climb rate, $c(h)$, is in feet per minute, and the new function, $m(h)$, is measured in minutes per foot, it follows that $m(h)$ is the reciprocal of $c(h)$, so $m(h) = \frac{1}{c(h)}$. Hence we get the following updated table.

h (feet)	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10,000
c (ft/min)	925	875	830	780	730	685	635	585	535	490	440
m (min/ft)	$\frac{1}{925}$	$\frac{1}{875}$	$\frac{1}{830}$	$\frac{1}{780}$	$\frac{1}{730}$	$\frac{1}{685}$	$\frac{1}{635}$	$\frac{1}{585}$	$\frac{1}{535}$	$\frac{1}{490}$	$\frac{1}{440}$

- Given an original function, we know that the units on its derivative are “units of input per unit of output.” Since the units on $m(h)$ are “minutes per foot”, it follows that the antiderivative of $m(h)$ should have output units of “minutes” since its input units are “feet”. The antiderivative function should thus tell us the total number of minutes that it takes for the plane to climb to an altitude of h feet.
- The number of minutes required for the airplane to ascend to 10,000 feet of altitude is given by the definite integral

$$M = \int_0^{10000} m(h) dh$$

since the units on $m(h)$ are “minutes per foot” and the units on Δh in the corresponding Riemann sum are “feet”, so when we take the product of quantities with these units, the resulting numerical quantity is measured in minutes.

- Reasoning similarly to our work in (c), the number of minutes required for the airplane to ascend to h feet of altitude is given by the definite integral

$$M(h) = \int_0^t m(t) dt.$$

Note that because h is the upper limit of integration, we need to use a different variable (t) of integration.

- e. The best possible estimates are provided by using the midpoint rule (we will learn why in Chapter 6, but this can also be determined by using left and right rules and thinking about the error for an always decreasing function like $m(h)$).

On the interval $[0, 6000]$, if we evaluate the function at every other data point, we can estimate the desired integral using 3 subintervals and find that

$$\begin{aligned}
 \int_0^{6000} m(h) dh &\approx M_3 \\
 &= m(1000) \cdot 2000 + m(3000) \cdot 2000 + m(5000) \cdot 2000 \\
 &= \left(\frac{1}{875} + \frac{1}{780} + \frac{1}{685} \right) \cdot 2000 \\
 &\approx 0.00388476 \cdot 2000 \\
 &\approx 7.77.
 \end{aligned}$$

Thus, it takes the plane a bit less than 8 minutes to climb to an altitude of 6,000 feet. Using 5 subintervals,

$$\begin{aligned}
 \int_0^{10000} m(h) dh &\approx M_5 \\
 &= m(1000) \cdot 2000 + m(3000) \cdot 2000 + m(5000) \cdot 2000 \\
 &\quad + m(7000) \cdot 2000 + m(9000) \cdot 2000 \\
 &= \left(\frac{1}{875} + \frac{1}{780} + \frac{1}{685} + \frac{1}{585} + \frac{1}{490} \right) \cdot 2000 \\
 &\approx 0.007634980475 \cdot 2000 \\
 &\approx 15.27.
 \end{aligned}$$

Thus, it takes the plane a bit more than 15 minutes to climb to an altitude of 10,000 feet. Note that in both of these computations, we are using M_n to denote the midpoint rule with n subintervals, and this “ M ” is different from the one we used earlier in $M(h)$.

5.5 Integration by Substitution

Motivating Questions

- How can we begin to find algebraic formulas for antiderivatives of more complicated algebraic functions?
 - What is an indefinite integral and how is its notation used in discussing antiderivatives?
 - How does the technique of u -substitution work to help us evaluate certain indefinite integrals, and how does this process rely on identifying function-derivative pairs?
-

In Section 4.4, we learned the key role that antiderivatives play in the process of evaluating definite integrals exactly. The Fundamental Theorem of Calculus tells us 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 5.2.2. Thus, if we wish to evaluate an integral such as

$$\int_0^1 (x^3 - \sqrt{x} + 5^x) dx,$$

it is straightforward to do so, since we can easily antiderivative $f(x) = x^3 - \sqrt{x} + 5^x$. Because one antiderivative of f is $F(x) = \frac{1}{4}x^4 - \frac{2}{3}x^{3/2} + \frac{1}{\ln(5)}5^x$, the Fundamental Theorem of Calculus tells us that

$$\begin{aligned} \int_0^1 (x^3 - \sqrt{x} + 5^x) dx &= \frac{1}{4}x^4 - \frac{2}{3}x^{3/2} + \frac{1}{\ln(5)}5^x \Big|_0^1 \\ &= \left(\frac{1}{4}(1)^4 - \frac{2}{3}(1)^{3/2} + \frac{1}{\ln(5)}5^1 \right) - \left(0 - 0 + \frac{1}{\ln(5)}5^0 \right) \\ &= -\frac{5}{12} + \frac{4}{\ln(5)}. \end{aligned}$$

We see that we have a natural interest in being able to find such algebraic antiderivatives. We emphasize *algebraic* antiderivatives, as opposed to any antiderivative, since we know by the Second 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. Our goal in this section is to “undo” the process of differentiation to find an algebraic antiderivative for a given function.

Example 5.5.1. In Section 2.5, we learned the Chain Rule and how it can be applied to find the derivative of a composite function. In particular, if u is a differentiable function of x , and f is a differentiable function of $u(x)$, then

$$\frac{d}{dx} [f(u(x))] = f'(u(x)) \cdot u'(x).$$

In words, we say that the derivative of a composite function $c(x) = f(u(x))$, 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.”

- a. For each of the following functions, use the Chain Rule to find the function’s derivative. Be sure to label each derivative by name (e.g., the derivative of $g(x)$ should be labeled $g'(x)$).

i. $g(x) = e^{3x}$

iv. $q(x) = (2 - 7x)^4$

ii. $h(x) = \sin(5x + 1)$

iii. $p(x) = \arctan(2x)$

v. $r(x) = 3^{4-11x}$

- b. For each of the following functions, use your work in (a) to help you determine the general antiderivative¹ of the function. Label each antiderivative by name (e.g., the antiderivative of m should be called M). In addition, check your work by computing the derivative of each proposed antiderivative.

i. $m(x) = e^{3x}$

iv. $v(x) = (2 - 7x)^3$

ii. $n(x) = \cos(5x + 1)$

iii. $s(x) = \frac{1}{1+4x^2}$

v. $w(x) = 3^{4-11x}$

- c. Based on your experience in parts (a) and (b), conjecture an antiderivative for each of the following functions. Test your conjectures by computing the derivative of each proposed antiderivative.

i. $a(x) = \cos(\pi x)$

iii. $c(x) = xe^{x^2}$

ii. $b(x) = (4x + 7)^{11}$

5.5.1 Reversing the Chain Rule: First Steps

Whenever f is a familiar function whose antiderivative is known and $u(x)$ is a linear function, it is straightforward to antidifferentiate a function of the form

$$h(x) = f(u(x)).$$

¹Recall that the general antiderivative of a function includes “ $+C$ ” to reflect the entire family of functions that share the same derivative.

Example 5.5.2. Determine the general antiderivative of

$$h(x) = (5x - 3)^6.$$

Check the result by differentiating.

For this composite function, the outer function f is $f(u) = u^6$, while the inner function is $u(x) = 5x - 3$. Since the antiderivative of f is $F(u) = \frac{1}{7}u^7 + C$, we see that the antiderivative of h is

$$H(x) = \frac{1}{7}(5x - 3)^7 \cdot \frac{1}{5} + C = \frac{1}{35}(5x - 3)^7 + C.$$

The inclusion of the constant $\frac{1}{5}$ is essential precisely because the derivative of the inner function is $u'(x) = 5$. Indeed, if we now compute $H'(x)$, we find by the Chain Rule (and Constant Multiple Rule) that

$$H'(x) = \frac{1}{35} \cdot 7(5x - 3)^6 \cdot 5 = (5x - 3)^6 = h(x),$$

and thus H is indeed the general antiderivative of h .

Hence, in the special case where the outer function is familiar and the inner function is linear, we can antidifferentiate composite functions according to the following rule.

If $h(x) = f(ax + b)$ and F is a known algebraic antiderivative of f , then the general antiderivative of h is given by

$$H(x) = \frac{1}{a}F(ax + b) + C.$$

It is useful to have shorthand notation that indicates the instruction to find an antiderivative. Thus, in a similar way to how the notation

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

represents the derivative of $f(x)$ with respect to x , we use the notation of the *indefinite integral*,

$$\int f(x) dx$$

to represent the general antiderivative of f with respect to x . Returning to the earlier example with $h(x) = (5x - 3)^6$, we can rephrase the relationship between h and its antiderivative H through the notation

$$\int (5x - 3)^6 dx = \frac{1}{35}(5x - 6)^7 + C.$$

When we find an antiderivative, we will often say that we *evaluate an indefinite integral*. Just as the notation $\frac{d}{dx}[\square]$ means “find the derivative with respect to x of \square ,” the notation $\int \square dx$ means “find a function of x whose derivative is \square .”

Example 5.5.3. Evaluate each of the following indefinite integrals. Check each antiderivative that you find by differentiating.

- | | |
|------------------------------|--|
| a. $\int \sin(8 - 3x) dx$ | d. $\int \csc(2x + 1) \cot(2x + 1) dx$ |
| b. $\int \sec^2(4x) dx$ | e. $\int \frac{1}{\sqrt{1-16x^2}} dx$ |
| c. $\int \frac{1}{11x-9} dx$ | f. $\int 5^{-x} dx$ |

Hint.

- a. Think $\int \sin(u) du$.
- b. Think $\int \sec^2(u) du$.
- c. Think $\int \frac{1}{u} du$.
- d. Think $\int \csc(u) \cot(u) du$.
- e. Think $\int \frac{1}{\sqrt{1-u^2}} du$.
- f. Think $\int 5^u du$.

Answer.

- a. $\int \sin(8 - 3x) dx = -\frac{1}{3}(-\cos(8 - 3x)) + C$.
- b. $\int \sec^2(4x) dx = \frac{1}{4} \tan(4x) + C$.
- c. $\int \frac{1}{11x-9} dx = \frac{1}{11} \ln(11x - 9) + C$.
- d. $\int \csc(2x + 1) \cot(2x + 1) dx = -\frac{1}{2} \cot(2x + 1) + C$.
- e. $\int \frac{1}{\sqrt{1-16x^2}} dx = \frac{1}{4} \arcsin(4x) + C$
- f. $\int 5^{-x} dx = -\frac{1}{\ln(5)} 5^{-x} + C$.

Solution.

- a. Since $u = 8 - 3x$ is linear and $\int \sin(u) du = -\cos(u) + C$, it follows that $\int \sin(8 - 3x) dx = -\frac{1}{3}(-\cos(8 - 3x)) + C$.
- b. Since $4x$ is linear and $\int \sec^2(u) du = \tan(u) + C$, we see $\int \sec^2(4x) dx = \frac{1}{4} \tan(4x) + C$.
- c. Using the fact that $\int \frac{1}{u} du = \ln|u| + C$, we have $\int \frac{1}{11x-9} dx = \frac{1}{11} \ln(11x - 9) + C$.
- d. We know $\int \csc(u) \cot(u) du = -\cot(u) + C$, so $\int \csc(2x + 1) \cot(2x + 1) dx = -\frac{1}{2} \cot(2x + 1) + C$.
- e. Observe that $\int \frac{1}{\sqrt{1-u^2}} dx = \arcsin(u) + C$, and thus viewing $16x^2 = (4x)^2$, we see that $\int \frac{1}{\sqrt{1-16x^2}} dx = \frac{1}{4} \arcsin(4x) + C$
- f. Since $\int 5^u du = \frac{1}{\ln(5)} 5^u + C$, we have that $\int 5^{-x} dx = -\frac{1}{\ln(5)} 5^{-x} + C$.

5.5.2 Reversing the Chain Rule: u -substitution

A natural question arises from our recent work: what happens when the inner function is not linear? For example, can we find antiderivatives of such functions as

$$g(x) = xe^{x^2} \text{ and } h(x) = e^{x^2}?$$

It is important to remember that differentiation and antiderivatives are almost inverse processes (that they are not is due to the $+C$ that arises when antiderivating). This almost-inverse relationship enables us to take any known derivative rule and rewrite it as a corresponding rule for an indefinite integral. For example, since

$$\frac{d}{dx} [x^5] = 5x^4,$$

we can equivalently write

$$\int 5x^4 dx = x^5 + C.$$

Recall that the Chain Rule states that

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

Restating this relationship in terms of an indefinite integral,

$$\int f'(g(x))g'(x) dx = f(g(x)) + C. \quad (5.5.1)$$

Equation (5.5.1) tells us that if we can view a given function as $f'(g(x))g'(x)$ for some appropriate choices of f and g , then we can antiderivatives the function by reversing the Chain Rule. Note 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* ($g(x)$ and $g'(x)$) when trying to apply the rule in Equation (5.5.1).

If we can identify a function-derivative pair, we will introduce a new variable u to represent the function $g(x)$. With $u = g(x)$, it follows in Leibniz notation that $\frac{du}{dx} = g'(x)$, so that in terms of differentials², $du = g'(x) dx$. Now converting the indefinite integral to a new one in terms of u , we have

$$\int f'(g(x))g'(x) dx = \int f'(u) du.$$

Provided that 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*, and summarize the rule as follows:

²If we recall from the definition of the derivative that $\frac{du}{dx} \approx \frac{\Delta u}{\Delta x}$ and use the fact that $\frac{du}{dx} = g'(x)$, then we see that $g'(x) \approx \frac{\Delta u}{\Delta x}$. Solving for Δu , $\Delta u \approx g'(x)\Delta x$. 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.

With the substitution $u = g(x)$,

$$\int f'(g(x))g'(x) dx = \int f'(u) du = f(u) + C = f(g(x)) + C.$$

To see u -substitution at work, we consider the following example.

Example 5.5.4. Evaluate the indefinite integral

$$\int x^3 \cdot \sin(7x^4 + 3) dx$$

and check the result by differentiating.

Solution. We can make two algebraic observations regarding the integrand, $x^3 \cdot \sin(7x^4 + 3)$. First, $\sin(7x^4 + 3)$ is a composite function; as such, we know we'll need a more sophisticated approach to antiderivatives. Second, x^3 is almost the derivative of $(7x^4 + 3)$; the only issue is a missing constant. Thus, x^3 and $(7x^4 + 3)$ are nearly 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.

Letting u represent the inner function of the composite function $\sin(7x^4 + 3)$, we have $u = 7x^4 + 3$, and thus $\frac{du}{dx} = 28x^3$. In differential notation, it follows that $du = 28x^3 dx$, and thus $x^3 dx = \frac{1}{28} du$. The original indefinite integral may be slightly rewritten as

$$\int \sin(7x^4 + 3) \cdot x^3 dx,$$

and so by substituting u for $7x^4 + 3$ and $\frac{1}{28} du$ for $x^3 dx$, it follows that

$$\int \sin(7x^4 + 3) \cdot x^3 dx = \int \sin(u) \cdot \frac{1}{28} du.$$

Now we may evaluate the easier integral in u , and then replace u by the expression $7x^4 + 3$. Doing so, we find

$$\begin{aligned} \int \sin(7x^4 + 3) \cdot x^3 dx &= \int \sin(u) \cdot \frac{1}{28} du \\ &= \frac{1}{28} \int \sin(u) du \\ &= \frac{1}{28}(-\cos(u)) + C \\ &= -\frac{1}{28} \cos(7x^4 + 3) + C. \end{aligned}$$

To check our work, we observe by the Chain Rule that

$$\frac{d}{dx} \left[-\frac{1}{28} \cos(7x^4 + 3) \right] = -\frac{1}{28} \cdot (-1) \sin(7x^4 + 3) \cdot 28x^3 = \sin(7x^4 + 3) \cdot x^3,$$

which is indeed the original integrand.

The u -substitution worked because the function multiplying $\sin(7x^4 + 3)$ was x^3 . If instead that function was x^2 or x^4 , the substitution process would not have worked. This is one of the primary challenges of antiderivatives: 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

$$\begin{aligned}\int xe^{x^2} dx &= \int e^u \cdot \frac{1}{2} du \\ &= \frac{1}{2} \int e^u du \\ &= \frac{1}{2} e^u + C \\ &= \frac{1}{2} e^{x^2} + C.\end{aligned}$$

However, for the similar indefinite integral

$$\int e^{x^2} dx,$$

the u -substitution $u = x^2$ is no longer possible because the factor of x is missing. 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 is the result of applying the Chain Rule to a different, related function.

Example 5.5.5. 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. $\int \frac{x^2}{5x^3+1} dx$

b. $\int e^x \sin(e^x) dx$

c. $\int \frac{\cos(\sqrt{x})}{\sqrt{x}} dx$

Hint.

- Note that $5x^3 + 1$ and $15x^2$ form a function-derivative pair.
- Recall that e^x is its own derivative.
- Observe that $x^{-1/2} = \frac{1}{\sqrt{x}}$.

Answer.

a. $\int \frac{x^2}{5x^3+1} dx = \frac{1}{15} \ln(5x^3 + 1) + C.$

b. $\int e^x \sin(e^x) dx = -\cos(e^x) + C.$

c. $\int \frac{\cos(\sqrt{x})}{\sqrt{x}} dx = 2 \sin(\sqrt{x}) + C.$

Solution.

- a. Since $5x^3 + 1$ and $15x^2$ form a function-derivative pair, we let $u = 5x^3 + 1$, and observe that $du = 15x^2 dx$, and thus $x^2 dx = \frac{1}{15} du$. Applying this substitution, integrating, and substituting back, $\int \frac{x^2}{5x^3+1} dx = \int \frac{\frac{1}{15} du}{u} = \frac{1}{15} \ln(u) + C = \frac{1}{15} \ln(5x^3 + 1) + C.$
- b. Because $\frac{d}{dx}[e^x] = e^x$, if we let $u = e^x$, it follows $du = e^x dx$. Substituting and integrating, $\int e^x \sin(e^x) dx = \int \sin(u) du = -\cos(u) + C = -\cos(e^x) + C.$
- c. Let $u = \sqrt{x}$, so that $du = \frac{1}{2}x^{-1/2} dx = \frac{dx}{2\sqrt{x}}$. We observe that $\frac{dx}{\sqrt{x}} = 2du$, and thus $\int \frac{\cos(\sqrt{x})}{\sqrt{x}} dx = \int 2 \cos(u) du = 2 \sin(u) + C = 2 \sin(\sqrt{x}) + C.$

5.5.3 Evaluating Definite Integrals via u -substitution

We have introduced u -substitution as a means to evaluate indefinite integrals of functions that can be written, up to a constant multiple, in the form $f(g(x))g'(x)$. This same technique can be used to evaluate definite integrals involving such functions, though 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

$$\begin{aligned} \int_{x=2}^{x=5} xe^{x^2} dx &= \int_{u=4}^{u=25} e^u \cdot \frac{1}{2} du \\ &= \frac{1}{2} e^u \Big|_{u=4}^{u=25} \end{aligned}$$

$$= \frac{1}{2}e^{25} - \frac{1}{2}e^4.$$

Alternatively, we could consider the related indefinite integral $\int xe^{x^2} dx$, find the antiderivative $\frac{1}{2}e^{x^2}$ through u -substitution, and then evaluate the original definite integral. With that method, 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.

Example 5.5.6. Evaluate each of the following definite integrals exactly through an appropriate u -substitution.

a. $\int_1^2 \frac{x}{1+4x^2} dx$

c. $\int_{2/\pi}^{4/\pi} \frac{\cos(\frac{1}{x})}{x^2} dx$

b. $\int_0^1 e^{-x}(2e^{-x} + 3)^9 dx$

Hint.

a. Let $u = 1 + 4x^2$.

b. $(2e^{-x} + 3)$ and e^{-x} form a function-derivative pair

c. $\frac{d}{dx} \left[\frac{1}{x} \right] = -\frac{1}{x^2}$

Answer.

a. $\int_{x=1}^{x=2} \frac{x}{1+4x^2} dx = \frac{1}{8}(\ln(17) - \ln(5)).$

b. $\int_0^1 e^{-x}(2e^{-x} + 3)^9 dx = -\frac{1}{20}(2e^{-1} + 3)^{10} + \frac{1}{20}(2e^0 + 3)^{10}.$

c. $\int_{2/\pi}^{4/\pi} \frac{\cos(\frac{1}{x})}{x^2} dx = 1 - \frac{\sqrt{2}}{2}.$

Solution.

a. Let $u = 1 + 4x^2$, so $du = 8x dx$ and $x dx = \frac{1}{8}du$. Note that $x = 1$ implies $u = 5$ and $x = 2$ implies $u = 17$. Thus, $\int_{x=1}^{x=2} \frac{x}{1+4x^2} dx = \frac{1}{8} \int_{u=5}^{u=17} \frac{du}{u} = \frac{1}{8} \ln(u) \Big|_5^{17} = \frac{1}{8}(\ln(17) - \ln(5)).$

b. First consider the corresponding indefinite integral, $\int e^{-x}(2e^{-x} + 3)^9 dx$, and let $u = 2e^{-x} + 3$ so that $du = -2e^{-x} dx$. We see $e^{-x} dx = -\frac{1}{2}du$, and thus $\int e^{-x}(2e^{-x} + 3)^9 dx = -\frac{1}{2} \int u^9 du = -\frac{1}{2} \cdot \frac{1}{10}u^{10} + C = -\frac{1}{20}(2e^{-x} + 3)^{10}$. Applying this antiderivative to the definite integral, we see that $\int_0^1 e^{-x}(2e^{-x} + 3)^9 dx = -\frac{1}{20}(2e^{-x} + 3)^{10} \Big|_0^1 = -\frac{1}{20}(2e^{-1} + 3)^{10} + \frac{1}{20}(2e^0 + 3)^{10}.$

c. Using the substitution $u = \frac{1}{x}$, it follows that $\int \frac{\cos(\frac{1}{x})}{x^2} dx = -\int \cos(u) du = -\sin(u) + C$. Hence, $\int_{2/\pi}^{4/\pi} \frac{\cos(\frac{1}{x})}{x^2} dx = -\sin\left(\frac{1}{x}\right) \Big|_{2/\pi}^{4/\pi} = -\sin\left(\frac{\pi}{4}\right) + \sin\left(\frac{\pi}{2}\right) = 1 - \frac{\sqrt{2}}{2}.$

5.5.4 Summary

- 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 to 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 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.

5.5.5 Exercises

1. (Product involving 4th power of a polynomial) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Product involving $\sin(x^6)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Fraction involving \ln^9) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Fraction involving e^{5x}) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Fraction involving $e^{5\sqrt{y}}$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Definite integral involving $e^{-\cos(q)}$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. This problem centers on finding antiderivatives for the basic trigonometric functions other than $\sin(x)$ and $\cos(x)$.
 - a. Consider the indefinite integral $\int \tan(x) dx$. By rewriting the integrand as $\tan(x) = \frac{\sin(x)}{\cos(x)}$ and identifying an appropriate function-derivative pair, make a u -substitution and hence evaluate $\int \tan(x) dx$.

- b. In a similar way, evaluate $\int \cot(x) dx$.
 c. Consider the indefinite integral

$$\int \frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} dx.$$

Evaluate this integral using the substitution $u = \sec(x) + \tan(x)$.

- d. Simplify the integrand in (c) by factoring the numerator. What is a far simpler way to write the integrand?
 e. Combine your work in (c) and (d) to determine $\int \sec(x) dx$.
 f. Using (c)-(e) as a guide, evaluate $\int \csc(x) dx$.

Answer.

- a. $\int \tan(x) dx = \ln(|\sec(x)|) + C$.
 b. $\int \cot(x) dx = -\ln(|\csc(x)|) + C$.
 c. $\int \frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} dx = \ln(|\sec(x) + \tan(x)|) + C$.
 d. $\frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} = \sec(x)$.
 e. $\int \sec(x) dx = \ln(|\sec(x) + \tan(x)|) + C$.
 f. $\int \csc(x) dx = -\ln(|\csc(x) + \cot(x)|) + C$.

Solution.

- a. We let $u = \cos(x)$, so $du = -\sin(x) dx$. This change of variable transforms the integral into one we can integrate as follows:

$$\begin{aligned} \int \tan(x) dx &= \int \frac{\sin(x)}{\cos(x)} dx \\ &= - \int \frac{1}{u} du \\ &= -\ln(|u|) + C \\ &= -\ln(|\cos(x)|) + C \\ &= \ln(|\cos(x)|^{-1}) + C \\ &= \ln(|\sec(x)|) + C. \end{aligned}$$

Note that we used the property of the logarithm that $p \ln(z) = \ln(z^p)$ as well as the fact that $\frac{1}{\cos(x)} = \sec(x)$ in order to simplify the result.

- b. We let $u = \sin(x)$, so $du = \cos(x) dx$. This change of variable transforms the integral into one we can integrate as follows:

$$\begin{aligned} \int \cot(x) dx &= \int \frac{\cos(x)}{\sin(x)} dx \\ &= \int \frac{1}{u} du \\ &= \ln(|u|) + C \end{aligned}$$

$$\begin{aligned}
 &= \ln(|\sin(x)|) + C \\
 &= -\ln(|\sin(x)|^{-1}) + C \\
 &= -\ln(|\csc(x)|) + C.
 \end{aligned}$$

Note that here we used the fact that $\ln(z) = (-1)(-1)\ln(z) = (-1)\ln(z^{-1})$ in order to simplify the result.

- c. We let $u = \sec(x) + \tan(x)$, so it follows that $du = (\sec(x)\tan(x) + \sec^2(x)) dx$. This change of variable shows that the basic structure of the integrand is simple and can be integrated as follows:

$$\begin{aligned}
 \int \frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} dx &= \int \frac{1}{u} du \\
 &= \ln(|u|) + C \\
 &= \ln(|\sec(x) + \tan(x)|) + C.
 \end{aligned}$$

- d. The numerator of $\frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)}$ can be rewritten by removing the common factor of $\sec(x)$ to obtain

$$\sec(x)(\sec(x) + \tan(x)).$$

The integrand can then be simplified to

$$\frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} = \frac{\sec(x)(\sec(x) + \tan(x))}{\sec(x) + \tan(x)} = \sec(x).$$

- e. Using the results of (c) and (d), we see that

$$\int \sec(x) dx = \int \frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} dx = \ln(|\sec(x) + \tan(x)|) + C.$$

- f. Following (c)-(e) we can evaluate $\int \csc(x) dx$ as follows. First, rewrite $\csc(x)$ by multiplying by $1 = \frac{\csc(x) + \cot(x)}{\csc(x) + \cot(x)}$, so that

$$\int \csc(x) dx = \int \frac{\csc(x)(\csc(x) + \cot(x))}{\csc(x) + \cot(x)} dx = \int \frac{\csc^2(x) + \csc(x)\cot(x)}{\csc(x) + \cot(x)} dx.$$

Now we evaluate the last integral in the preceding equation with the substitution $u = \csc(x) + \cot(x)$. Then $du = -(\csc(x)\cot(x) + \csc^2(x)) dx$. This change of variable transforms the integral so that we can integrate:

$$\begin{aligned}
 \int \frac{\csc^2(x) + \csc(x)\cot(x)}{\csc(x) + \cot(x)} dx &= - \int \frac{1}{u} du \\
 &= -\ln(|u|) + C \\
 &= -\ln(|\csc(x) + \cot(x)|) + C.
 \end{aligned}$$

Connecting the equalities we have established in our work, we have shown that

$$\int \csc(x) dx = -\ln(|\csc(x) + \cot(x)|) + C.$$

8. Consider the indefinite integral $\int x\sqrt{x-1} dx$.

- At first glance, this integrand may not seem suited to substitution due to the presence of x in separate locations in the integrand. Nonetheless, using the composite function $\sqrt{x-1}$ as a guide, let $u = x - 1$. Determine expressions for both x and dx in terms of u .
- Convert the given integral in x to a new integral in u .
- Evaluate the integral in (b) by noting that $\sqrt{u} = u^{1/2}$ and observing that it is now possible to rewrite the integrand in u by expanding through multiplication.
- Evaluate each of the integrals $\int x^2\sqrt{x-1} dx$ and $\int x\sqrt{x^2-1} dx$. Write a paragraph to discuss the similarities among the three indefinite integrals in this problem and the role of substitution and algebraic rearrangement in each.

Answer.

- $\int x\sqrt{x-1} dx = \int (u+1)\sqrt{u} du$.
- $\int x\sqrt{x-1} dx = \frac{2}{5}(x-1)^{5/2} + \frac{2}{3}(x-1)^{3/2} + C$.
- $\int x^2\sqrt{x-1} dx = \frac{2}{7}(x-1)^{7/2} + \frac{4}{5}(x-1)^{5/2} + \frac{2}{3}(x-1)^{3/2} + C$.
- $\int x\sqrt{x^2-1} dx = \frac{1}{3}(x^2-1)^{3/2} + C$.

Solution.

- Letting $u = x - 1$, we find that $x = u + 1$ and thus $dx = du$.
- Using the change of variables in (a),

$$\int x\sqrt{x-1} dx = \int (u+1)\sqrt{u} du.$$

- Using the fact that $\sqrt{u} = u^{1/2}$ and distributing, we can evaluate the integral as follows:

$$\begin{aligned} \int x\sqrt{x-1} dx &= \int (u+1)u^{1/2} du \\ &= \int (u^{3/2} + u^{1/2}) du \\ &= \frac{2}{5}u^{5/2} + \frac{2}{3}u^{3/2} + C \\ &= \frac{2}{5}(x-1)^{5/2} + \frac{2}{3}(x-1)^{3/2} + C. \end{aligned}$$

- First, for $\int x^2\sqrt{x-1} dx$, we let $u = x - 1$ so that $x = u + 1$ and $dx = du$. From there, our work is very similar to what we did in (b):

$$\begin{aligned} \int x^2\sqrt{x-1} dx &= \int (u+1)^2 u^{1/2} du \\ &= \int (u^2 + 2u + 1) u^{1/2} du \\ &= \int (u^{5/2} + 2u^{3/2} + u^{1/2}) du \end{aligned}$$

$$\begin{aligned}
 &= \frac{2}{7}u^{7/2} + \frac{4}{5}u^{5/2} + \frac{2}{3}u^{3/2} + C \\
 &= \frac{2}{7}(x-1)^{7/2} + \frac{4}{5}(x-1)^{5/2} + \frac{2}{3}(x-1)^{3/2} + C.
 \end{aligned}$$

For $\int x\sqrt{x^2-1} dx$, despite the fact that the integral looks remarkably similar, the situation is quite different. We observe that with $u = x^2 - 1$, it follows $du = 2x dx$, so we have a standard function-derivative pair. Noting that $x dx = \frac{1}{2} du$, it follows that

$$\begin{aligned}
 \int x\sqrt{x^2-1} dx &= \int \sqrt{u} \cdot \frac{1}{2} du \\
 &= \frac{1}{2} \cdot \frac{2}{3}u^{3/2} + C \\
 &= \frac{1}{3}u^{3/2} + C \\
 &= \frac{1}{3}(x^2-1)^{3/2} + C.
 \end{aligned}$$

In the situation where we have $(x-a)$ under a square root and a polynomial outside the square root, the substitution $u = x-a$ will allow us to convert the integrand into an expression we can integrate by changing the form of the radical from $\sqrt{x-a}$ to \sqrt{u} . From there we are able to distribute and integrate an expression that is essentially a sum of (fractional) powers of u . In the situation where a polynomial of degree higher than 1 (such as x^2-1) is under the radical, we likely have to be fortunate enough to have the derivative of said polynomial (up to a constant) outside the radical in order to be able to integrate, as we were in the last example.

9. Consider the indefinite integral $\int \sin^3(x) dx$.
 - a. Explain why the substitution $u = \sin(x)$ will not work to help evaluate the given integral.
 - b. Recall the Fundamental Trigonometric Identity, which states that $\sin^2(x) + \cos^2(x) = 1$. By observing that $\sin^3(x) = \sin(x) \cdot \sin^2(x)$, use the Fundamental Trigonometric Identity to rewrite the integrand as the product of $\sin(x)$ with another function.
 - c. Explain why the substitution $u = \cos(x)$ now provides a possible way to evaluate the integral in (b).
 - d. Use your work in (a)-(c) to evaluate the indefinite integral $\int \sin^3(x) dx$.
 - e. Use a similar approach to evaluate $\int \cos^3(x) dx$.

Answer.

- a. We don't have a function-derivative pair.
- b. $\sin^3(x) = \sin(x)(1 - \cos^2(x))$.
- c. $u = \cos(x)$ and $du = -\sin(x) dx$.
- d. $\int \sin^3(x) dx = \frac{1}{3} \cos^3(x) - \cos(x) + C$.
- e. $\int \cos^3(x) dx = \sin(x) - \frac{1}{3} \sin^3(x) + C$.

Solution.

- a. If we let $u = \sin(x)$, then $du = \cos(x) dx$. Since there is no cosine function present in the integrand, we don't have the needed function-derivative pair present in order to transform the integral into one we can evaluate.
- b. Using the Fundamental Trigonometric Identity, we observe that $\sin^2(x) = 1 - \cos^2(x)$, and thus

$$\sin^3(x) = \sin(x) \cdot \sin^2(x) = \sin(x)(1 - \cos^2(x)).$$

- c. By our work in (b), we see that $\int \sin^3(x) dx = \int \sin(x)(1 - \cos^2(x)) dx$, and in this integral we have the function-derivative pair that corresponds to $u = \cos(x)$ and $du = -\sin(x) dx$.
- d. Using this substitution and evaluating the resulting integral,

$$\begin{aligned} \int \sin^3(x) dx &= \int \sin(x)(1 - \cos^2(x)) dx \\ &= \int (-1)(1 - u^2) du \\ &= \int (u^2 - 1) du \\ &= \frac{1}{3}u^3 - u + C \\ &= \frac{1}{3}\cos^3(x) - \cos(x) + C. \end{aligned}$$

- e. Here we observe that $\cos^3(x) = \cos(x) \cdot \cos^2(x) = \cos(x)(1 - \sin^2(x))$, so letting $u = \sin(x)$ and thus $du = \cos(x) dx$, we have

$$\begin{aligned} \int \cos^3(x) dx &= \int \cos(x)(1 - \sin^2(x)) dx \\ &= \int (1 - u^2) du \\ &= u - \frac{1}{3}u^3 + C \\ &= \sin(x) - \frac{1}{3}\sin^3(x) + C. \end{aligned}$$

- 10.** 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 megawatts³ at time t . Units are critical throughout this problem.

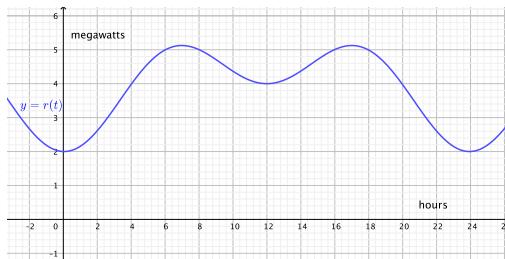
- a. Sketch a carefully labeled graph of $r(t)$ on the interval $[0, 24]$ and explain its meaning. Why is this a reasonable model of power consumption?

³The unit *megawatt* is itself a rate, which measures energy consumption per unit time. A *megawatt-hour* is the total amount of energy that is equivalent to a constant stream of 1 megawatt of power being sustained for 1 hour.

- 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 energy Mathland consumes in a typical day.
- d. What is Mathland's average rate of power consumption in a given 24-hour period? What are the units on this quantity?

Answer.

- a. The model is reasonable because it appears to be periodic and the rate of consumption seems to peak at the times of day where people are most active in their homes.

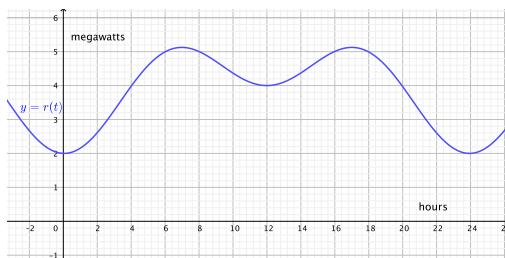


- b. The total power consumed in 24 hours, measured in megawatt-hours.

- c. $\int_0^{24} r(t) dt \approx 95.7809$ megawatts of power used in 24 hours.
- d. $r_{\text{AVG}}[0,24] \approx 3.99087$ megawatts.

Solution.

- a. In the following plot, we see that the model is reasonable because it appears to be periodic and repeats every 24 hours, as well as that the rate of consumption seems to peak at the times of day where people are most active in their homes: around 7 a.m. and around 5 p.m.



- b. The meaning of $\int_0^{24} r(t) dt$ is the total power consumed in 24 hours, measured in megawatt-hours. This is because $r(t)$ is measured in megawatts and Δt in hours, and thus in the Riemann sum that corresponds to the definite integral, the terms in the sum of the form $r(t) \cdot \Delta t$ have units "megawatt-hours". Moreover, by integrating the rate of power consumption we get the total change in power consumed.
- c. In a typical day, we know that Mathland consumes $\int_0^{24} r(t) dt$ megawatt-hours of power, and thus we seek to evaluate the integral

$$\int_0^{24} r(t) dt = \int_0^{24} [4 + \sin(0.263t + 4.7) + \cos(0.526t + 9.4)] dt.$$

To do so, we use the substitutions $u = 0.263t + 4.7$ (so $du = 0.263 dt$) and $w = 0.526t + 9.4$ (so $dw = 0.526 dt$) to help us antidifferentiate the two trigonometric expressions, and thus we find that

$$\int \sin(0.263t + 4.7) dt = \frac{1}{0.263} \int \sin(u) du = -\frac{1}{0.263} \cos(u) = -\frac{1}{0.263} \cos(0.263t + 4.7)$$

and

$$\int \cos(0.526t + 9.4) dt = \frac{1}{0.526} \int \cos(w) dw = \frac{1}{0.526} \sin(w) = \frac{1}{0.526} \sin(0.526t + 9.4).$$

Putting this all together and using the Fundamental Theorem of Calculus, we have

$$\begin{aligned} \int_0^{24} r(t) dt &= 4t - \frac{1}{0.263} \cos(0.263t + 4.7) + \frac{1}{0.526} \sin(0.526t + 9.4) \Big|_0^{24} \\ &= \left[4 \cdot 24 - \frac{1}{0.263} \cos(0.263 \cdot 24 + 4.7) + \frac{1}{0.526} \sin(0.526 \cdot 24 + 9.4) \right] - \left[-\frac{1}{0.263} \cos(4.7) + \frac{1}{0.526} \sin(9.4) \right] \\ &\approx 95.7809 \end{aligned}$$

megawatts of power used in 24 hours.

- d. We know that the average value of the rate function r on $[0, 24]$ is

$$r_{\text{AVG}[0,24]} = \frac{1}{24-0} \int_0^{24} r(t) dt.$$

Thus, by our work in (c), we have that

$$r_{\text{AVG}[0,24]} \approx \frac{1}{24} \cdot 95.7809 = 3.99087$$

which is measured in megawatts.

5.6 Integration by Parts

Motivating Questions

- How do we evaluate indefinite integrals that involve products of basic functions such as $\int x \sin(x) dx$ and $\int xe^x dx$?
 - What is the method of integration by parts and how can we consistently apply it to integrate products of basic functions?
 - How does the algebraic structure of functions guide us in identifying u and dv in using integration by parts?
-

In Section 5.5, we learned the technique of u -substitution for evaluating indefinite integrals. For example, the indefinite integral $\int x^3 \sin(x^4) dx$ is perfectly suited to u -substitution, because one factor is a composite function and the other factor is the derivative (up to a constant) of the inner function. Recognizing the algebraic structure of a function can help us to find its antiderivative.

Next we consider integrands with a different elementary algebraic structure: a product of basic functions. For instance, suppose we are interested in evaluating the indefinite integral

$$\int x \sin(x) dx.$$

The integrand is the product of the basic functions $f(x) = x$ and $g(x) = \sin(x)$. We know that it is relatively complicated to compute the derivative of the product of two functions, so we should expect that antiderivativing a product should be similarly involved. Intuitively, we expect that evaluating $\int x \sin(x) dx$ will involve somehow reversing the Product Rule.

To that end, in Example 5.6.1 we refresh our understanding of the Product Rule and then investigate some indefinite integrals that involve products of basic functions.

Example 5.6.1. In Section 2.2, we developed the Product Rule and studied how it is employed to differentiate a product of two functions. In particular, recall that if f and g are differentiable functions of x , then

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

- a. For each of the following functions, use the Product Rule to find the function's derivative. Be sure to label each derivative by name (e.g., the derivative of $g(x)$ should be labeled $g'(x)$).

- | | |
|------------------------|--------------------------|
| i. $g(x) = x \sin(x)$ | iv. $q(x) = x^2 \cos(x)$ |
| ii. $h(x) = xe^x$ | |
| iii. $p(x) = x \ln(x)$ | v. $r(x) = e^x \sin(x)$ |

- b. Use your work in (a) to help you evaluate the following indefinite integrals. Use differentiation to check your work.

- | | |
|--|--------------------------------------|
| i. $\int xe^x + e^x \, dx$ | iv. $\int x \cos(x) + \sin(x) \, dx$ |
| ii. $\int e^x(\sin(x) + \cos(x)) \, dx$ | |
| iii. $\int 2x \cos(x) - x^2 \sin(x) \, dx$ | v. $\int 1 + \ln(x) \, dx$ |

- c. Observe that the examples in (b) work nicely because of the derivatives you were asked to calculate in (a). Each integrand in (b) is precisely the result of differentiating one of the products of basic functions found in (a). To see what happens when an integrand is still a product but not necessarily the result of differentiating an elementary product, we consider how to evaluate

$$\int x \cos(x) \, dx.$$

- i. First, observe that

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

Integrating both sides indefinitely and using the fact that the integral of a sum is the sum of the integrals, we find that

$$\int \left(\frac{d}{dx} [x \sin(x)] \right) \, dx = \int x \cos(x) \, dx + \int \sin(x) \, dx.$$

In this last equation, evaluate the indefinite integral on the left side as well as the rightmost indefinite integral on the right.

- ii. In the most recent equation from (i.), solve the equation for the expression $\int x \cos(x) \, dx$.
 iii. For which product of basic functions have you now found the antiderivative?

5.6.1 Reversing the Product Rule: Integration by Parts

Problem (c) in Example 5.6.1 provides a clue to the general technique known as Integration by Parts, which comes from reversing the Product Rule. Recall that the Product Rule states that

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

Integrating both sides of this equation indefinitely with respect to x , we find

$$\int \frac{d}{dx} [f(x)g(x)] \, dx = \int f(x)g'(x) \, dx + \int g(x)f'(x) \, dx. \quad (5.6.1)$$

On the left side of Equation (5.6.1), we have the indefinite integral of the derivative of a function. Temporarily omitting the constant that may arise, we have

$$f(x)g(x) = \int f(x)g'(x) dx + \int g(x)f'(x) dx. \quad (5.6.2)$$

We solve for the first indefinite integral on the left to generate the rule

$$\int f(x)g'(x) dx = f(x)g(x) - \int g(x)f'(x) dx. \quad (5.6.3)$$

Often we express Equation (5.6.3) in terms of the variables u and v , where $u = f(x)$ and $v = g(x)$. In differential notation, $du = f'(x) dx$ and $dv = g'(x) dx$, so we can state the rule for Integration by Parts in its most common form as follows:

$$\int u dv = uv - \int v du.$$

To apply integration by parts, we look for a product of basic functions that we can identify as u and dv . If we can antiderivative dv to find v , and evaluating $\int v du$ is not more difficult than evaluating $\int u dv$, then this substitution usually proves to be fruitful. To demonstrate, we consider the following example.

Example 5.6.2. Evaluate the indefinite integral

$$\int x \cos(x) dx$$

using integration by parts.

Solution. When we use integration by parts, we have a choice for u and dv . In this problem, we can either let $u = x$ and $dv = \cos(x) dx$, or let $u = \cos(x)$ and $dv = x dx$. While there is not a universal rule for how to choose u and dv , a good guideline is this: do so in a way that $\int v du$ is at least as simple as the original problem $\int u dv$.

This leads us to choose¹ $u = x$ and $dv = \cos(x) dx$, from which it follows that $du = 1 dx$ and $v = \sin(x)$. With this substitution, the rule for integration by parts tells us that

$$\int x \cos(x) dx = x \sin(x) - \int \sin(x) \cdot 1 dx.$$

All that remains to do is evaluate the (simpler) integral $\int \sin(x) \cdot 1 dx$. Doing so, we find

$$\int x \cos(x) dx = x \sin(x) - (-\cos(x)) + C = x \sin(x) + \cos(x) + C.$$

Observe that when we get to the final stage of evaluating the last remaining antiderivative, it is at this step that we include the integration constant, $+C$.

¹Observe that if we considered the alternate choice, and let $u = \cos(x)$ and $dv = x dx$, then $du = -\sin(x) dx$ and $v = \frac{1}{2}x^2$, from which we would write $\int x \cos(x) dx = \frac{1}{2}x^2 \cos(x) - \int \frac{1}{2}x^2(-\sin(x)) dx$. Thus we have replaced the problem of integrating $x \cos(x)$ with that of integrating $\frac{1}{2}x^2 \sin(x)$; the latter is clearly more complicated, which shows that this alternate choice is not as helpful as the first choice.

The general technique of integration by parts involves trading the problem of integrating the product of two functions for the problem of integrating the product of two related functions. That is, we convert the problem of evaluating $\int u \, dv$ to that of evaluating $\int v \, du$. This clearly shapes our choice of u and v . In Example 5.6.2, the original integral to evaluate was $\int x \cos(x) \, dx$, and through the substitution provided by integration by parts, we were instead able to evaluate $\int \sin(x) \cdot 1 \, dx$. Note that the original function x was replaced by its derivative, while $\cos(x)$ was replaced by its antiderivative.

Example 5.6.3. Evaluate each of the following indefinite integrals. Check each antiderivative that you find by differentiating.

- | | |
|-----------------------------|-----------------------------|
| a. $\int te^{-t} \, dt$ | c. $\int z \sec^2(z) \, dz$ |
| b. $\int 4x \sin(3x) \, dx$ | d. $\int x \ln(x) \, dx$ |

Hint.

- a. Try $u = t$.
- b. Let $dv = \sin(3x)dx$
- c. Remember that $\int \sec^2(z) \, dz = \tan(z)$.
- d. Note that $\ln(x)dx$ has a simple derivative to work with.

Answer.

- a. $\int te^{-t} \, dt = -te^{-t} - e^{-t}$.
- b. $\int 4x \sin(3x) \, dx = -\frac{4}{3}x \cos(3x) + \frac{4}{9} \sin(3x) + c$.
- c. $\int z \sec^2(z) \, dz = z \tan(z) + \ln(\cos(z)) + c$.
- d. $\int x \ln(x) \, dx = \frac{1}{2}x^2 \ln(x) - \frac{1}{4}x^2 + c$.

Solution.

- a. Using $u = t$ and $dv = e^{-t} \, dt$, we obtain $du = dt$ and $v = -e^{-t}$. So

$$\begin{aligned} \int te^{-t} \, dt &= -te^{-t} + \int e^{-t} \, dt \\ &= -te^{-t} - e^{-t} + c \end{aligned}$$

We check using differentiation as follows:

$$\begin{aligned} \frac{d}{dt} (-te^{-t} - e^{-t} + c) &= (te^{-t} - e^{-t}) + e^{-t} \\ &= te^{-t} \end{aligned}$$

- b. Using $u = 4x$ and $dv = \sin(3x)dx$, we obtain $du = 4dx$ and $v = -\frac{1}{3} \cos(3x)$. So

$$\begin{aligned}\int 4x \sin(3x)dx &= -\frac{4}{3}x \cos(3x) + \frac{4}{3} \int \cos(3x)dx \\ &= -\frac{4}{3}x \cos(3x) + \frac{4}{9} \sin(3x) + c\end{aligned}$$

We check using differentiation as follows:

$$\begin{aligned}\frac{d}{dx} \left(-\frac{4}{3}x \cos(3x) + \frac{4}{9} \sin(3x) + c \right) &= \left(4x \sin(3x) - \frac{4}{3} \cos(3x) \right) + \frac{4}{3} \cos(3x) \\ &= 4x \sin(3x)\end{aligned}$$

- c. Using $u = z$ and $dv = \sec^2(z)dz$, we obtain $du = dz$ and $v = \tan(z)$. So

$$\begin{aligned}\int z \sec^2(z)dz &= z \tan(z) - \int \tan(z)dz \\ &= z \tan(z) + \ln(\cos(z)) + c\end{aligned}$$

We check using differentiation as follows:

$$\begin{aligned}\frac{d}{dz} (z \tan(z) + \ln(\cos(z)) + c) &= z \sec^2(z) + \tan(z) - \frac{\sin(z)}{\cos(z)} \\ &= z \sec^2(z) + \tan(z) - \tan(z) \\ &= z \sec^2(z)\end{aligned}$$

- d. Using $u = \ln(x)$ and $dv = xdx$, we obtain $du = \frac{1}{x}dx$ and $v = \frac{1}{2}x^2$. So

$$\begin{aligned}\int x \ln(x)dx &= \frac{1}{2}x^2 \ln(x) - \frac{1}{2} \int xdx \\ &= \frac{1}{2}x^2 \ln(x) - \frac{1}{4}x^2 + c\end{aligned}$$

We check using differentiation as follows:

$$\begin{aligned}\frac{d}{dx} \left(\frac{1}{2}x^2 \ln(x) - \frac{1}{4}x^2 + c \right) &= \frac{1}{2}x + x \ln(x) - \frac{1}{2}x \\ &= x \ln(x)\end{aligned}$$

5.6.2 Some Subtleties with Integration by Parts

Sometimes integration by parts is not an obvious choice, but the technique is appropriate nonetheless. Integration by parts allows us to replace one function in a product with its derivative while replacing the other with its antiderivative. For instance, consider evaluating

$$\int \arctan(x) dx.$$

Initially, this problem seems ill-suited to integration by parts, since there does not appear to be a product of functions present. But if we note that $\arctan(x) = \arctan(x) \cdot 1$, and realize that we know the derivative of $\arctan(x)$ as well as the antiderivative of 1, we see the possibility for the substitution $u = \arctan(x)$ and $dv = 1 \, dx$. We explore this substitution further in Example 5.6.4.

In a related problem, consider $\int t^3 \sin(t^2) \, dt$. Observe that there is a composite function present in $\sin(t^2)$, but there is not an obvious function-derivative pair, as we have t^3 (rather than simply t) multiplying $\sin(t^2)$. In this problem we use both u -substitution and integration by parts. First we write $t^3 = t \cdot t^2$ and consider the indefinite integral

$$\int t \cdot t^2 \cdot \sin(t^2) \, dt.$$

We let $z = t^2$ so that $dz = 2t \, dt$, and thus $t \, dt = \frac{1}{2} \, dz$. (We are using the variable z to perform a “ z -substitution” first so that we may then apply integration by parts.) Under this z -substitution, we now have

$$\int t \cdot t^2 \cdot \sin(t^2) \, dt = \int z \cdot \sin(z) \cdot \frac{1}{2} \, dz.$$

The resulting integral can be evaluated by parts. This, too, is explored further in Example 5.6.4.

These problems show that we sometimes must think creatively in choosing the variables for substitution in integration by parts, and that we may need to use substitution for an additional change of variables.

Example 5.6.4. Evaluate each of the following indefinite integrals, using the provided hints.

- Evaluate $\int \arctan(x) \, dx$ by using Integration by Parts with the substitution $u = \arctan(x)$ and $dv = 1 \, dx$.
- Evaluate $\int \ln(z) \, dz$. Consider a similar substitution to the one in (a).
- Use the substitution $z = t^2$ to transform the integral $\int t^3 \sin(t^2) \, dt$ to a new integral in the variable z , and evaluate that new integral by parts.
- Evaluate $\int s^5 e^{s^3} \, ds$ using an approach similar to that described in (c).
- Evaluate $\int e^{2t} \cos(e^t) \, dt$. You will find it helpful to note that $e^{2t} = e^t \cdot e^t$.

Answer.

- $\int \arctan(x) \, dx = x \arctan(x) - \frac{1}{2} \ln(1 + x^2) + c$.
- $\int \ln(z) \, dz = z \ln(z) - z + c$.
- $\int t^3 \sin(t^2) \, dt = \frac{1}{2}(-t^2 \cos(t^2) + \sin(t^2))$.
- $\int s^5 e^{s^3} \, ds = \frac{1}{3}(s^3 e^{s^3} - e^{s^3}) + c$.

e. $\int e^{2t} \cos(e^t) dt = e^t \sin(e^t) + \cos(e^t) + c.$

Solution.

a. We use integration by parts with $u = \arctan(x)$ and $dv = dx$. So $du = \frac{1}{1+x^2} dx$ and $v = x$.

$$\begin{aligned}\int \arctan(x) dx &= uv - \int v du \\ &= x \arctan(x) - \int \frac{x}{1+x^2} dx\end{aligned}$$

For the integral in the last equation, use the substitution $z = 1 + x^2$ and $dz = 2x dx$.

$$\begin{aligned}\int \arctan(x) dx &= x \arctan(x) - \int \frac{x}{1+x^2} dx \\ &= x \arctan(x) - \frac{1}{2} \int \frac{1}{z} dz \\ &= x \arctan(x) - \frac{1}{2} \ln(z) + c \\ &= x \arctan(x) - \frac{1}{2} \ln(1+x^2) + c\end{aligned}$$

b. Using $u = \ln(z)$ and $dv = dz$, we obtain $du = \frac{1}{z} dz$ and $v = z$. So

$$\begin{aligned}\int \ln(z) dz &= z \ln(z) - \int dz \\ &= z \ln(z) - z + c\end{aligned}$$

c. Using $z = t^2$, we obtain $dz = 2t dt$ and $\frac{1}{2} dz = t dt$. So

$$\begin{aligned}\int t^3 \sin(t^2) dt &= \int t^2 \sin(t^2) \cdot t dt \\ &= \int z \sin(z) \frac{1}{2} dz \\ &= \frac{1}{2} \int z \sin(z) dz\end{aligned}$$

Now using integration by parts with $u = z$ and $dv = \sin(z) dz$, we obtain $du = dz$ and $v = -\cos(z)$. So

$$\begin{aligned}\int t^3 \sin(t^2) dt &= \frac{1}{2} \int z \sin(z) dz \\ &= \frac{1}{2} \left(-z \cos(z) + \int \cos(z) dz \right) \\ &= \frac{1}{2} (-z \cos(z) + \sin(z))\end{aligned}$$

Finally, we use $z = t^2$ and obtain

$$\int t^3 \sin(t^2) dt = \frac{1}{2} (-t^2 \cos(t^2) + \sin(t^2)).$$

d. Use a substitution with $z = s^3$ and $dz = 3s^2 ds$.

$$\begin{aligned} \int s^5 e^{s^3} ds &= \int s^2 s^3 e^{s^3} ds \\ &= \frac{1}{3} \int z e^z dz \end{aligned}$$

Now use integration by parts with $u = z$ and $dv = e^z dz$. So $du = dz$ and $v = e^z$.

$$\begin{aligned} \int s^5 e^{s^3} ds &= \frac{1}{3} \int z e^z dz \\ &= \frac{1}{3} \left(z e^z - \int e^z dz \right) \\ &= \frac{1}{3} (z e^z - e^z) + c \\ &= \frac{1}{3} (s^3 e^{s^3} - e^{s^3}) + c \end{aligned}$$

e. Using $z = e^t$, we see that $dz = e^t dt$ and so

$$\begin{aligned} \int e^{2t} \cos(e^t) dt &= \int e^t \cos(e^t) \cdot e^t dt \\ &= \int z \cos(z) dz \end{aligned}$$

In Example 5.6.2, we saw that

$$\int z \cos(z) dz = z \sin(z) + \cos(z) + c.$$

Combining these two results, we obtain

$$\begin{aligned} \int e^{2t} \cos(e^t) dt &= \int z \cos(z) dz \\ &= z \sin(z) + \cos(z) + c \\ &= e^t \sin(e^t) + \cos(e^t) + c \end{aligned}$$

5.6.3 Using Integration by Parts Multiple Times

Integration by parts is well suited to integrating the product of basic functions, allowing us to trade a given integrand for a new one where one function in the product is replaced by its derivative, and the other is replaced by its antiderivative. The goal in this trade of $\int u \, dv$ for $\int v \, du$ is that the new integral be simpler to evaluate than the original one. Sometimes it is necessary to apply integration by parts more than once in order to evaluate a given integral.

Example 5.6.5. Evaluate $\int t^2 e^t dt$.

Solution. Let $u = t^2$ and $dv = e^t dt$. Then $du = 2t dt$ and $v = e^t$, and thus

$$\int t^2 e^t dt = t^2 e^t - \int 2t e^t dt.$$

The integral on the right side is simpler to evaluate than the one on the left, but it still requires integration by parts. Now letting $u = 2t$ and $dv = e^t dt$, we have $du = 2 dt$ and $v = e^t$, so that

$$\int t^2 e^t dt = t^2 e^t - \left(2t e^t - \int 2e^t dt \right).$$

(Note the parentheses, which remind us to distribute the minus sign to the entire value of the integral $\int 2t e^t dt$.) The final integral on the right is a basic one; evaluating that integral and distributing the minus sign, we find

$$\int t^2 e^t dt = t^2 e^t - 2t e^t + 2e^t + C.$$

Of course, even more than two applications of integration by parts may be necessary. In the preceding example, if the integrand had been $t^3 e^t$, we would have had to use integration by parts three times.

Next, we consider the slightly different scenario.

Example 5.6.6. Evaluate $\int e^t \cos(t) dt$.

Solution. We can choose to let u be either e^t or $\cos(t)$; we pick $u = \cos(t)$, and thus $dv = e^t dt$. With $du = -\sin(t) dt$ and $v = e^t$, integration by parts tells us that

$$\int e^t \cos(t) dt = e^t \cos(t) - \int e^t (-\sin(t)) dt,$$

or equivalently that

$$\int e^t \cos(t) dt = e^t \cos(t) + \int e^t \sin(t) dt. \quad (5.6.4)$$

The new integral has the same algebraic structure as the original one. While the overall situation isn't necessarily better than what we started with, it hasn't gotten worse. Thus, we proceed to integrate by parts again. This time we let $u = \sin(t)$ and $dv = e^t dt$, so that $du = \cos(t) dt$ and $v = e^t$, which implies

$$\int e^t \cos(t) dt = e^t \cos(t) + \left(e^t \sin(t) - \int e^t \cos(t) dt \right). \quad (5.6.5)$$

We seem to be back where we started, as two applications of integration by parts has led us back to the original problem, $\int e^t \cos(t) dt$. But if we look closely at Equation (5.6.5), we see that we can use algebra to solve for the value of the desired integral. Adding $\int e^t \cos(t) dt$ to both sides of the equation, we have

$$2 \int e^t \cos(t) dt = e^t \cos(t) + e^t \sin(t),$$

and therefore

$$\int e^t \cos(t) dt = \frac{1}{2} (e^t \cos(t) + e^t \sin(t)) + C.$$

Note that since we never actually encountered an integral we could evaluate directly, we didn't have the opportunity to add the integration constant C until the final step.

Example 5.6.7. Evaluate each of the following indefinite integrals.

a. $\int x^2 \sin(x) dx$

b. $\int t^3 \ln(t) dt$

c. $\int e^z \sin(z) dz$

d. $\int s^2 e^{3s} ds$

e. $\int t \arctan(t) dt$ (*Hint:* At a certain point in this problem, it is very helpful to note that $\frac{t^2}{1+t^2} = 1 - \frac{1}{1+t^2}$.)

Hint.

a. Start with $u = x^2$.

b. Begin with $u = t^3$.

c. You'll have to integrate by parts twice.

d. Start with $u = s^2$.

e. Try $u = \arctan(x)$.

Answer.

a. $\int x^2 \sin(x) dx = -x^2 \cos(x) + 2x \sin(x) + 2 \cos(x) + C.$

b. $\int t^3 \ln(t) dt = \frac{1}{4} t^4 \ln(t) - \frac{1}{16} t^4 + C.$

c. $\int e^z \sin(z) dz = -\frac{1}{2} e^z \cos(z) + \frac{1}{2} e^z \sin(z) + C.$

d. $\int s^2 e^{3s} ds = \frac{1}{3} s^2 e^{3s} - \frac{2}{9} s e^{3s} + \frac{2}{27} e^{3s} + C.$

e. $\int t \arctan(t) dt = \frac{1}{2} t^2 \arctan(t) - \frac{1}{2} t - \frac{1}{2} \arctan(t) + C.$

Solution.

a. First, using $u = x^2$ and $dv = \sin(x)dx$, we get $du = 2x dx$ and $v = -\cos(x)$. So

$$\int x^2 \sin(x) dx = -x^2 \cos(x) + 2 \int x \cos(x) dx.$$

Now, for $\int x \cos(x)dx$, we use $u = x$ and $dv = \cos(x)$ and get $du = dx$ and $v = \sin(x)$. So

$$\begin{aligned}\int x^2 \sin(x)dx &= -x^2 \cos(x) + 2 \left(x \sin(x) - \int \sin(x)dx \right) \\ &= -x^2 \cos(x) + 2x \sin(x) + 2 \cos(x) + c\end{aligned}$$

b. Using $u = \ln(t)$ and $dv = t^3 dt$, we get $du = \frac{1}{t} dt$ and $v = \frac{1}{4}t^4$. So

$$\begin{aligned}\int t^3 \ln(t)dt &= \frac{1}{4}t^4 \ln(t) - \frac{1}{4} \int t^3 dt \\ &= \frac{1}{4}t^4 \ln(t) - \frac{1}{4} \cdot \frac{1}{4}t^4 + c \\ &= \frac{1}{4}t^4 \ln(t) - \frac{1}{16}t^4 + c\end{aligned}$$

c. We use $u = e^z$ and $dv = \sin(z)dz$. This gives $du = e^z dz$ and $v = -\cos(z)$, and

$$\int e^z \sin(z)dz = -e^z \cos(z) + \int e^z \cos(z)dz.$$

For $\int e^z \cos(z)dz$, we use $u = e^z$ and $dv = \cos(z)dz$. So $du = e^z dz$ and $v = \sin(z)$, and

$$\begin{aligned}\int e^z \sin(z)dz &= -e^z \cos(z) + e^z \sin(z)dz - \int e^z \sin(z)dz \\ 2 \int e^z \sin(z)dz &= -e^z \cos(z) + e^z \sin(z) + c' \\ \int e^z \sin(z)dz &= -\frac{1}{2}e^z \cos(z) + \frac{1}{2}e^z \sin(z) + c'\end{aligned}$$

d. We use $u = s^2$, $dv = e^{3s}ds$, $du = 2sds$, and $v = \frac{1}{3}e^{3s}$. This gives

$$\int s^2 e^{3s} ds = \frac{1}{3}s^2 e^{3s} - \frac{2}{3} \int s e^{3s} ds.$$

We now use $u = s$, $dv = e^{3s}ds$, $du = ds$, and $v = \frac{1}{3}e^{3s}$ for $\int s e^{3s} ds$.

$$\begin{aligned}\int s^2 e^{3s} ds &= \frac{1}{3}s^2 e^{3s} - \frac{2}{3} \left(\frac{1}{3}se^{3s} - \frac{1}{3} \int e^{3s} ds \right) \\ &= \frac{1}{3}s^2 e^{3s} - \frac{2}{3} \left(\frac{1}{3}se^{3s} - \frac{1}{9}e^{3s} \right) + c \\ &= \frac{1}{3}s^2 e^{3s} - \frac{2}{9}se^{3s} + \frac{2}{27}e^{3s} + c\end{aligned}$$

e. Using $u = \arctan(t)$, $dv = tdt$, $du = \frac{1}{1+t^2}dt$, and $v = \frac{1}{2}t^2$, we get

$$\int t \arctan(t)dt = \frac{1}{2}t^2 \arctan(t) - \frac{1}{2} \int \frac{t^2}{1+t^2} dt$$

$$\begin{aligned}
 &= \frac{1}{2}t^2 \arctan(t) - \frac{1}{2} \int \left(1 - \frac{1}{1+t^2}\right) dt \\
 &= \frac{1}{2}t^2 \arctan(t) - \frac{1}{2}t - \frac{1}{2} \arctan(t) + c
 \end{aligned}$$

5.6.4 Evaluating Definite Integrals Using Integration by Parts

We can use the technique of integration by parts to evaluate a definite integral.

Example 5.6.8. Evaluate

$$\int_0^{\pi/2} t \sin(t) dt.$$

Solution. One option is to find an antiderivative (using indefinite integral notation) and then apply the Fundamental Theorem of Calculus to find that

$$\begin{aligned}
 \int_0^{\pi/2} t \sin(t) dt &= (-t \cos(t) + \sin(t)) \Big|_0^{\pi/2} \\
 &= \left(-\frac{\pi}{2} \cos\left(\frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2}\right)\right) - (-0 \cos(0) + \sin(0)) \\
 &= 1.
 \end{aligned}$$

Alternatively, we can apply integration by parts and work with definite integrals throughout. With this method, we must remember to evaluate the product uv over the given limits of integration. Using the substitution $u = t$ and $dv = \sin(t) dt$, so that $du = dt$ and $v = -\cos(t)$, we write

$$\begin{aligned}
 \int_0^{\pi/2} t \sin(t) dt &= -t \cos(t) \Big|_0^{\pi/2} - \int_0^{\pi/2} (-\cos(t)) dt \\
 &= -t \cos(t) \Big|_0^{\pi/2} + \sin(t) \Big|_0^{\pi/2} \\
 &= \left(-\frac{\pi}{2} \cos\left(\frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2}\right)\right) - (-0 \cos(0) + \sin(0)) \\
 &= 1.
 \end{aligned}$$

As with any substitution technique, it is important to use notation carefully and completely, and to ensure that the end result makes sense.

5.6.5 When u -substitution and Integration by Parts Fail to Help

Both integration techniques we have discussed apply in relatively limited circumstances. It is not hard to find examples of functions for which neither technique produces an antiderivative; indeed, there are many, many functions that appear elementary but that do not have an elementary algebraic antiderivative. For instance, neither u -substitution nor integration by parts proves fruitful for the indefinite integrals

$$\int e^{x^2} dx \text{ and } \int x \tan(x) dx.$$

While there are other integration techniques, some of which we will consider briefly, none of them enables us to find an algebraic antiderivative for e^{x^2} or $x \tan(x)$. We do know from the Second Fundamental Theorem of Calculus that we can construct an integral antiderivative for each function; $F(x) = \int_0^x e^{t^2} dt$ is an antiderivative of $f(x) = e^{x^2}$, and $G(x) = \int_0^x t \tan(t) dt$ is an antiderivative of $g(x) = x \tan(x)$. But finding an elementary algebraic formula that doesn't involve integrals for either F or G turns out not only to be impossible through u -substitution or integration by parts, but indeed impossible altogether. Antidifferentiation is much harder in general than differentiation.

5.6.6 Summary

- Through the method of integration by parts, we can evaluate indefinite integrals that involve products of basic functions such as $\int x \sin(x) dx$ and $\int x \ln(x) dx$. Using a substitution enables us to trade one of the functions in the product for its derivative, and the other for its antiderivative, in an effort to find a different product of functions that is easier to integrate.
- If the algebraic structure of an integrand is a product of basic functions in the form $\int f(x)g'(x) dx$, we can use the substitution $u = f(x)$ and $dv = g'(x) dx$ and apply the rule

$$\int u \, dv = uv - \int v \, du$$

to evaluate the original integral $\int f(x)g'(x) dx$ by instead evaluating

$$\int v \, du = \int f'(x)g(x) dx.$$

- When deciding to integrate by parts, we have to select both u and dv . That selection is guided by the overall principle that the new integral $\int v \, du$ not be more difficult than the original integral $\int u \, dv$. In addition, it is often helpful to recognize if one of the functions present is much easier to differentiate than antidifferentiate (such as $\ln(x)$), in which case that function often is best assigned the variable u . In addition, dv must be a function that we can antidifferentiate.

5.6.7 Exercises

1. (Choose which method to use) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Product involving $\cos(5x)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Product involving e^{8z}) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Definite integral of te^{-t}) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

5. Let $f(t) = te^{-2t}$ and $F(x) = \int_0^x f(t) dt$.

- Determine $F'(x)$.
- Use the First FTC to find a formula for F that does not involve an integral.
- Is F an increasing or decreasing function for $x > 0$? Why?

Answer.

- $F'(x) = xe^{-2x}$.
- $F(x) = -\frac{1}{2}xe^{-2x} - \frac{1}{4}e^{-2x} + \frac{1}{4}$
- Increasing.

Solution.

- By the Second Fundamental Theorem of Calculus, $F'(x) = f(x) = xe^{-2x}$.
- To use the First Fundamental Theorem of Calculus to evaluate the integral in $F(x) = \int_0^x te^{-2t} dt$, we use integration by parts. Letting

$$\begin{aligned} u &= t, \quad du = dt \\ dv &= e^{-2t} dt, \quad v = -\frac{1}{2}e^{-2t}, \end{aligned}$$

it follows that

$$\int te^{-2t} dt = -\frac{1}{2}te^{-2t} - \int \left(-\frac{1}{2}e^{-2t}\right) dt = -\frac{1}{2}te^{-2t} - \frac{1}{4}e^{-2t}.$$

Thus,

$$\begin{aligned} F(x) &= \int_0^x te^{-2t} dt \\ &= -\frac{1}{2}te^{-2t} - \frac{1}{4}e^{-2t} \Big|_0^x \\ &= -\frac{1}{2}xe^{-2x} - \frac{1}{4}e^{-2x} - \left(0 - \frac{1}{4}e^0\right) \\ &= -\frac{1}{2}xe^{-2x} - \frac{1}{4}e^{-2x} + \frac{1}{4} \end{aligned}$$

- For $x > 0$, we observe that $F'(x) = f(x) = xe^{-2x} > 0$ since e to any power is positive. Thus, because $F'(x) > 0$, it follows that F is increasing for all $x > 0$.
- Consider the indefinite integral given by $\int e^{2x} \cos(e^x) dx$.
 - Noting that $e^{2x} = e^x \cdot e^x$, use the substitution $z = e^x$ to determine a new, equivalent integral in the variable z .
 - Evaluate the integral you found in (a) using an appropriate technique.
 - How is the problem of evaluating $\int e^{2x} \cos(e^{2x}) dx$ different from evaluating the integral in (a)? Do so.
 - Evaluate each of the following integrals as well, keeping in mind the approach(es) used earlier in this problem:

- $\int e^{2x} \sin(e^x) dx$
- $\int e^{3x} \sin(e^{3x}) dx$
- $\int x e^{x^2} \cos(e^{x^2}) \sin(e^{x^2}) dx$

Answer.

- $\int e^{2x} \cos(e^x) dx = \int z \cos(z) dz.$
- $\int e^{2x} \cos(e^x) dx = e^x \sin(e^x) + \cos(e^x) + C.$
- $\int e^{2x} \cos(e^{2x}) dx = \frac{1}{2} \sin(e^{2x}) + C$
- $\int e^{2x} \sin(e^x) dx = \sin(e^x) - z \cos(e^x) + C.$
 - $\int e^{3x} \sin(e^{3x}) dx = -\frac{1}{3} \cos(e^{3x}) + C.$
 - $\int x e^{x^2} \cos(e^{x^2}) \sin(e^{x^2}) dx = \frac{1}{4} \sin^2(e^{x^2}) + C.$

Solution.

- We first write

$$\int e^{2x} \cos(e^x) dx = \int e^x \cdot e^x \cos(e^x) dx.$$

Letting $z = e^x$, we have $dz = e^x dx$, and thus

$$\int e^{2x} \cos(e^x) dx = \int z \cos(z) dz.$$

- We can evaluate the integral $\int z \cos(z) dz$ using integration by parts in the standard way with $u = z$ and $dv = \cos(z) dz$. Doing so,

$$\int z \cos(z) dz = z \sin(z) - \int \sin(z) dz = z \sin(z) + \cos(z) + C.$$

Returning to the original integral in x and using the substitution we first used with z , we have shown that

$$\int e^{2x} \cos(e^x) dx = e^x \sin(e^x) + \cos(e^x) + C.$$

- For the integral $\int e^{2x} \cos(e^{2x}) dx$, we actually have a function-derivative pair (up to a constant multiple) already present: letting $u = e^{2x}$, $du = 2e^{2x} dx$, and thus we can evaluate the integral using this substitution. In particular,

$$\begin{aligned} \int e^{2x} \cos(e^{2x}) dx &= \int \cos(u) \cdot \frac{1}{2} du \\ &= \frac{1}{2} \sin(u) + C \\ &= \frac{1}{2} \sin(e^{2x}) + C \end{aligned}$$

- For the first integral, $\int e^{2x} \sin(e^x) dx$, since e^x is the inside function in the composition, we let $z = e^x$, so $dz = e^x dx$, and write $e^{2x} = e^x \cdot e^x$. Hence,

$$\int e^{2x} \sin(e^x) dx = \int e^x \cdot e^x \sin(e^x) dx = \int z \sin(z) dz.$$

Using the standard integration by parts with $u = z$ and $dv = \sin(z) dz$, we find

$$\int z \sin(z) dz = \sin(z) - z \cos(z) + C.$$

Re-substituting from z back to x , we've found that

$$\int e^{2x} \sin(e^x) dx = \sin(e^x) - z \cos(e^x) + C.$$

- The integral $\int e^{3x} \sin(e^{3x}) dx$ can be evaluated using a u -substitution since $u = e^{3x}$ is present to form a function-derivative pair with $du = 3e^{3x}$. It follows that

$$\int e^{3x} \sin(e^{3x}) dx = \int \sin(u) \cdot \frac{1}{3} du = -\frac{1}{3} \cos(u) + C = -\frac{1}{3} \cos(e^{3x}) + C.$$

- For the final integral, $\int x e^{x^2} \cos(e^{x^2}) \sin(e^{x^2}) dx$, we first observe by the chain rule that $\frac{d}{dx} [e^{x^2}] = 2xe^{x^2}$, which shows that we have a function-derivative pair present. Using the substitution $z = e^{x^2}$, it follows $dz = 2xe^{x^2} dx$, and therefore

$$\int x e^{x^2} \cos(e^{x^2}) \sin(e^{x^2}) dx = \int \cos(z) \sin(z) \cdot \frac{1}{2} dz.$$

The new integral in z can be evaluated using several different approaches, including u -substitution with $u = \sin(z)$, $du = \cos(z) dz$. Doing so,

$$\int \cos(z) \sin(z) \cdot \frac{1}{2} dz = \frac{1}{2} \int u du = \frac{1}{2} \cdot \frac{1}{2} u^2 + C = \frac{1}{4} \sin^2(z) + C.$$

Returning to the original integral in x and our earlier work, substituting backward we find that

$$\int x e^{x^2} \cos(e^{x^2}) \sin(e^{x^2}) dx = \frac{1}{4} \sin^2(e^{x^2}) + C.$$

7. For each of the following indefinite integrals, determine whether you would use u -substitution, integration by parts, neither*, or both to evaluate the integral. In each case, write one sentence to explain your reasoning, and include a statement of any substitutions used. (That is, if you decide in a problem to let $u = e^{3x}$, you should state that, as well as that $du = 3e^{3x} dx$.) Finally, use your chosen approach to evaluate each integral. (* one of the following problems does not have an elementary antiderivative and you are not expected to actually evaluate this integral; this will correspond with a choice of "neither" among those given.)

- $\int x^2 \cos(x^3) dx$
- $\int x^5 \cos(x^3) dx$ (Hint: $x^5 = x^2 \cdot x^3$)
- $\int x \ln(x^2) dx$
- $\int \sin(x^4) dx$
- $\int x^3 \sin(x^4) dx$
- $\int x^7 \sin(x^4) dx$

Answer.

- u -substitution; $\int x^2 \cos(x^3) dx = \frac{1}{3} \sin(x^3) + C$.
- Both are needed; $\int x^5 \cos(x^3) dx = \frac{1}{3} (x^3 \sin(x^3) + \cos(x^3)) + C$.
- Integration by parts; $\int x \ln(x^2) dx = \frac{x^2}{2} \ln(x^2) - \frac{x^2}{2} + C$.
- Neither.
- u -substitution; $\int x^3 \sin(x^4) dx = -\frac{1}{4} \cos(x^4) + C$.
- Both are needed; $\int x^7 \sin(x^4) dx = -\frac{1}{4} x^4 \cos(x^4) + \frac{1}{4} \sin(x^4) + C$.

Solution.

- We will use a u -substitution with $u = x^3$ because the integrand involves a composite function with x^3 as the innermost function and because $du = 3x^2 dx$ involves the factor x^2 which is also present in the integrand. When we make this substitution we will have $x^2 dx = \frac{1}{3} du$ and the integral can be evaluated as follows:

$$\begin{aligned}\int x^2 \cos(x^3) dx &= \frac{1}{3} \int \cos(u) du \\ &= \frac{1}{3} \sin(u) + C \\ &= \frac{1}{3} \sin(x^3) + C.\end{aligned}$$

- We might try substituting $z = x^3$, since there is a composite function in the integrand. In this case we have $dz = 3x^2 dx$. To make this substitution we need to rewrite the integrand as

$$x^5 \cos(x^3) = x^3 [x^2 \cos(x^3)]$$

and notice that $x^3 = z$. Then

$$\int x^5 \cos(x^3) dx = \frac{1}{3} \int z \cos(z) dz.$$

We can then try integration by parts on this last integral with

$$\begin{aligned}u &= z, \quad du = dz \\ dv &= \cos(z) dz, \quad v = \sin(z)\end{aligned}$$

We then have

$$\begin{aligned}\int x^5 \cos(x^3) dx &= \frac{1}{3} \int z \cos(z) dz \\ &= \frac{1}{3} \left(z \sin(z) - \int \sin(z) dz \right) \\ &= \frac{1}{3} (z \sin(z) + \cos(z)) + C \\ &= \frac{1}{3} (x^3 \sin(x^3) + \cos(x^3)) + C.\end{aligned}$$

- c. Since the logarithm function changes to a rational function when we differentiate it, using integration by parts with $u = \ln(x^2) = 2\ln(x)$ will make the resulting integral easier to evaluate. So we let

$$u = \ln(x^2), \quad du = \frac{2}{x} dx$$

$$dv = x dx, \quad v = \frac{x^2}{2}$$

and then integration by parts yields

$$\begin{aligned} \int x \ln(x^2) dx &= \frac{x^2}{2} \ln(x^2) - \int \left(\frac{2}{x}\right) \left(\frac{x^2}{2}\right) dx \\ &= \frac{x^2}{2} \ln(x^2) - \int x dx \\ &= \frac{x^2}{2} \ln(x^2) - \frac{x^2}{2} + C. \end{aligned}$$

- d. We might try a substitution with $u = x^4$, but then $du = 4x^3 dx$ and there is no factor of x^3 in the integrand to allow us to complete this substitution. If we attempt to integrate by parts, then only option is to let $u = \sin(x^4)$ which makes $du = 4x^3 \cos(x^4)$ and we only make the integral more complicated. This is one integral we cannot evaluate with the tools we have.
- e. We will try a u -substitution with $u = x^4$, because the integrand involves a composite function with x^4 as the innermost function and because $du = 4x^3 dx$ involves the factor x^3 which is also present in the integrand. When we make this substitution we have $x^3 dx = \frac{1}{4} du$ and the integral can be evaluated as follows:

$$\begin{aligned} \int x^3 \sin(x^4) dx &= \frac{1}{4} \int \sin(u) du \\ &= -\frac{1}{4} \cos(u) + C \\ &= -\frac{1}{4} \cos(x^4) + C. \end{aligned}$$

- f. We will use integration by parts. Notice that

$$\int x^7 \sin(x^4) dx = \int x^4 [x^3 \sin(x^4)] dx$$

and we have a factor we can differentiate (x^3) and one we can integrate ($x^3 \sin(x^4)$) (as we did in part (e)). So we let

$$u = x^3, \quad du = 3x^2 dx$$

$$dv = x^3 \sin(x^4) dx, \quad v = -\frac{1}{4} \cos(x^4)$$

Integration by parts formula now gives

$$\int x^7 \sin(x^4) dx = -\frac{1}{4} x^4 \cos(x^4) + \int x^3 \cos(x^4) dx.$$

This last integral is essentially the same as obtaining v from dv , so we will make the substitution $z = x^4$ with $dz = 4x^3 dx$. Changing variables in the last integral gives us

$$\begin{aligned}\int x^7 \sin(x^4) dx &= -\frac{1}{4}x^4 \cos(x^4) + \int x^3 \cos(x^4) dx \\ &= -\frac{1}{4}x^4 \cos(x^4) + \frac{1}{4} \int \cos(z) dz \\ &= -\frac{1}{4}x^4 \cos(x^4) + \frac{1}{4} \sin(z) + C \\ &= -\frac{1}{4}x^4 \cos(x^4) + \frac{1}{4} \sin(x^4) + C.\end{aligned}$$

5.7 Other Options for Finding Algebraic Antiderivatives

Motivating Questions

- How does the method of partial fractions enable any rational function to be antiderivatived?
- What role have integral tables historically played in the study of calculus and how can a table be used to evaluate integrals such as $\int \sqrt{a^2 + u^2} du$?
- What role can a computer algebra system play in the process of finding antiderivatives?

We have learned two antiderivative techniques: u -substitution and integration by parts. The former is used to reverse the chain rule, while the latter to reverse the product rule. But we have seen that each works only in specialized circumstances. For example, while $\int xe^{x^2} dx$ may be evaluated by u -substitution and $\int xe^x dx$ by integration by parts, neither method provides a route to evaluate $\int e^{x^2} dx$, and in fact an elementary algebraic antiderivative for e^{x^2} does not exist. No antiderivative method will provide us with a simple algebraic formula for a function $F(x)$ that satisfies $F'(x) = e^{x^2}$.

In this section of the text, our main goals are to identify some classes of functions that can be antiderivatived, and to learn some methods to do so. We should also recognize that there are many functions for which an algebraic formula for an antiderivative does not exist, and appreciate the role that computing technology can play in finding antiderivatives of other complicated functions.

Example 5.7.1. For each of the indefinite integrals below, the main question is to decide whether the integral can be evaluated using u -substitution, integration by parts, a combination of the two, or neither. For integrals for which your answer is affirmative, state the substitution(s) you would use. It is not necessary to actually evaluate any of the integrals completely, unless the integral can be evaluated immediately using a familiar basic antiderivative.

- $\int x^2 \sin(x^3) dx, \int x^2 \sin(x) dx, \int \sin(x^3) dx, \int x^5 \sin(x^3) dx$
- $\int \frac{1}{1+x^2} dx, \int \frac{x}{1+x^2} dx, \int \frac{2x+3}{1+x^2} dx, \int \frac{e^x}{1+(e^x)^2} dx,$
- $\int x \ln(x) dx, \int \frac{\ln(x)}{x} dx, \int \ln(1+x^2) dx, \int x \ln(1+x^2) dx,$
- $\int x \sqrt{1-x^2} dx, \int \frac{1}{\sqrt{1-x^2}} dx, \int \frac{x}{\sqrt{1-x^2}} dx, \int \frac{1}{x\sqrt{1-x^2}} dx,$

5.7.1 The Method of Partial Fractions

The method of partial fractions is used to integrate rational functions. It involves reversing the process of finding a common denominator.

Example 5.7.2. Evaluate

$$\int \frac{5x}{x^2 - x - 2} dx.$$

Solution. If we factor the denominator, we can see how R might be the sum of two fractions of the form $\frac{A}{x-2} + \frac{B}{x+1}$, so we suppose that

$$\frac{5x}{(x-2)(x+1)} = \frac{A}{x-2} + \frac{B}{x+1}$$

and look for the constants A and B .

Multiplying both sides of this equation by $(x-2)(x+1)$, we find that

$$5x = A(x+1) + B(x-2).$$

Since we want this equation to hold for every value of x , we can use insightful choices of specific x -values to help us find A and B . Taking $x = -1$, we have

$$5(-1) = A(0) + B(-3),$$

so $B = \frac{5}{3}$. Choosing $x = 2$, it follows

$$5(2) = A(3) + B(0),$$

so $A = \frac{10}{3}$. Thus,

$$\int \frac{5x}{x^2 - x - 2} dx = \int \frac{10/3}{x-2} + \frac{5/3}{x+1} dx.$$

This integral is straightforward to evaluate, and hence we find that

$$\int \frac{5x}{x^2 - x - 2} dx = \frac{10}{3} \ln|x-2| + \frac{5}{3} \ln|x+1| + C.$$

It turns out that we can use the method of partial fractions to rewrite any rational function $R(x) = \frac{P(x)}{Q(x)}$ where the degree of the polynomial P is less than¹ the degree of Q as a sum of simpler rational functions of one of the following forms:

$$\frac{A}{x-c}, \frac{A}{(x-c)^n}, \frac{Ax+B}{x^2+k}, \text{ or } \frac{Ax+B}{(x^2+k)^n}$$

where A , B , and c are real numbers, and k is a positive real number. Because we can antidifferentiate each of these basic forms, partial fractions enables us to antidifferentiate any rational function.

A computer algebra system such as *Maple*, *Mathematica*, or *WolframAlpha* can be used to find the partial fraction decomposition of any rational function. In *WolframAlpha*, entering

partial fraction $5x/(x^2-x-2)$

¹If the degree of P is greater than or equal to the degree of Q , long division may be used to write R as the sum of a polynomial plus a rational function where the numerator's degree is less than the denominator's.

results in the output

$$\frac{5x}{x^2 - x - 2} = \frac{10}{3(x - 2)} + \frac{5}{3(x + 1)}.$$

We will use technology to generate partial fraction decompositions of rational functions, and then evaluate the integrals using established methods.

Example 5.7.3. For each of the following problems, evaluate the integral by using the partial fraction decomposition provided.

- $\int \frac{1}{x^2 - 2x - 3} dx$, given that $\frac{1}{x^2 - 2x - 3} = \frac{1/4}{x-3} - \frac{1/4}{x+1}$
- $\int \frac{x^2+1}{x^3-x^2} dx$, given that $\frac{x^2+1}{x^3-x^2} = -\frac{1}{x} - \frac{1}{x^2} + \frac{2}{x-1}$
- $\int \frac{x-2}{x^4+x^2} dx$, given that $\frac{x-2}{x^4+x^2} = \frac{1}{x} - \frac{2}{x^2} + \frac{-x+2}{1+x^2}$

Hint.

- $\int \ln(u) du = \frac{1}{u} + C$.
- $\frac{1}{x^2} = x^{-2}$.
- $\frac{-x+2}{1+x^2} = \frac{-x}{1+x^2} + \frac{2}{1+x^2}$,

Answer.

- $\int \frac{1}{x^2 - 2x - 3} dx = \frac{1}{4} \ln|x - 3| - \frac{1}{4} \ln|x + 1| + C$.
- $\int \frac{x^2+1}{x^3-x^2} dx = -\ln(x) + x^{-1} + 2 \ln(x - 1) + C$.
- $\int \frac{x-2}{x^4+x^2} dx = \ln|x| + 2x^{-1} - \frac{1}{2} \ln(1+x^2) + 2 \arctan(x) + C$.

Solution.

- $\int \frac{1}{x^2 - 2x - 3} dx = \int \frac{1/4}{x-3} - \frac{1/4}{x+1} dx = \frac{1}{4} \ln|x - 3| - \frac{1}{4} \ln|x + 1| + C$.
- $\int \frac{x^2+1}{x^3-x^2} dx = \int \left(-\frac{1}{x} - \frac{1}{x^2} + \frac{2}{x-1}\right) dx = -\ln(x) + x^{-1} + 2 \ln(x - 1) + C$.
- By the partial fractions decomposition, $\int \frac{x-2}{x^4+x^2} dx = \int \frac{1}{x} - \frac{2}{x^2} + \frac{-x+2}{1+x^2} dx$. Observing that $\frac{-x+2}{1+x^2} = \frac{-x}{1+x^2} + \frac{2}{1+x^2}$, it now follows that

$$\int \frac{x-2}{x^4+x^2} dx = \int \frac{1}{x} - \frac{2}{x^2} - \frac{x}{1+x^2} + \frac{2}{1+x^2} dx.$$

Noting that $\int \frac{x}{1+x^2} dx$ can be evaluated by the u -substitution $u = 1 + x^2$, we see that $\int \frac{x}{1+x^2} dx = \frac{1}{2} \ln(1+x^2) + C$. Thus

$$\int \frac{x-2}{x^4+x^2} dx = \int \frac{1}{x} - \frac{2}{x^2} - \frac{x}{1+x^2} + \frac{2}{1+x^2} dx = \ln|x| + 2x^{-1} - \frac{1}{2} \ln(1+x^2) + 2 \arctan(x) + C.$$

5.7.2 Using an Integral Table

Calculus has a long history, going back to Greek mathematicians in 400-300 BC. Its main foundations were first investigated and understood independently by Isaac Newton and Gottfried Wilhelm Leibniz in the late 1600s, making the modern ideas of calculus well over 300 years old. It is instructive to realize that until the late 1980s, the personal computer did not exist, so calculus (and other mathematics) had to be done by hand for roughly 300 years. In the 21st century, however, computers have revolutionized many aspects of the world we live in, including mathematics. In this section we take a short historical tour to precede discussing the role computer algebra systems can play in evaluating indefinite integrals. In particular, we consider a class of integrals involving certain radical expressions.

As seen in the short table of integrals found in Appendix A, there are many forms of integrals that involve $\sqrt{a^2 \pm w^2}$ and $\sqrt{w^2 - a^2}$. These integral rules can be developed using a technique known as *trigonometric substitution* that we choose to omit; instead, we will simply accept the results presented in the table. To see how these rules are used, consider the differences among

$$\int \frac{1}{\sqrt{1-x^2}} dx, \quad \int \frac{x}{\sqrt{1-x^2}} dx, \quad \text{and} \quad \int \sqrt{1-x^2} dx.$$

The first integral is a familiar basic one, and results in $\arcsin(x) + C$. The second integral can be evaluated using a standard u -substitution with $u = 1 - x^2$. The third, however, is not familiar and does not lend itself to u -substitution.

In Appendix A, we find the rule

$$(h) \int \sqrt{a^2 - u^2} du = \frac{u}{2} \sqrt{a^2 - u^2} + \frac{a^2}{2} \arcsin \frac{u}{a} + C.$$

Using the substitutions $a = 1$ and $u = x$ (so that $du = dx$), it follows that

$$\int \sqrt{1-x^2} dx = \frac{x}{2} \sqrt{1-x^2} - \frac{1}{2} \arcsin x + C.$$

Whenever we are applying a rule in the table, we are doing a u -substitution, especially when the substitution is more complicated than setting $u = x$ as in the last example.

Example 5.7.4. Evaluate the integral

$$\int \sqrt{9+64x^2} dx.$$

Solution. Here, we want to use Rule (c) from the table, but we now set $a = 3$ and $u = 8x$. We also choose the “+” option in the rule. With this substitution, it follows that $du = 8dx$, so $dx = \frac{1}{8}du$. Applying the substitution,

$$\int \sqrt{9+64x^2} dx = \int \sqrt{9+u^2} \cdot \frac{1}{8} du = \frac{1}{8} \int \sqrt{9+u^2} du.$$

By Rule (c), we now find that

$$\begin{aligned}\int \sqrt{9+64x^2} dx &= \frac{1}{8} \left(\frac{u}{2} \sqrt{u^2 + 9} + \frac{9}{2} \ln |u + \sqrt{u^2 + 9}| + C \right) \\ &= \frac{1}{8} \left(\frac{8x}{2} \sqrt{64x^2 + 9} + \frac{9}{2} \ln |8x + \sqrt{64x^2 + 9}| + C \right).\end{aligned}$$

Whenever we use a u -substitution in conjunction with Appendix A, it's important that we not forget to address any constants that arise and include them in our computations, such as the $\frac{1}{8}$ that appeared in Example 5.7.4.

Example 5.7.5. For each of the following integrals, evaluate the integral using u -substitution and/or an entry from the table found in Appendix A.

a. $\int \sqrt{x^2 + 4} dx$

c. $\int \frac{2}{\sqrt{16+25x^2}} dx$

b. $\int \frac{x}{\sqrt{x^2+4}} dx$

d. $\int \frac{1}{x^2\sqrt{49-36x^2}} dx$

Hint.

a. Compare $\int \sqrt{u^2 + a^2} du$.

b. Try a straightforward u -substitution; the table is unneeded.

c. Let $a = 4$ and $u = 5x$ and look for a similar integral in the table.

d. Let $a = 7$ and $u = 6x$; find a related integral in the table.

Answer.

a. $\int \sqrt{x^2 + 4} dx = \frac{1}{2} \arctan\left(\frac{x}{2}\right) + C.$

b. $\int \frac{x}{\sqrt{x^2+4}} dx = \sqrt{x^2 + 4} + C.$

c. $\int \frac{2}{\sqrt{16+25x^2}} dx = \frac{2}{5} \ln |5x + \sqrt{16 + 25x^2}| + C.$

d. $\int \frac{1}{x^2\sqrt{49-36x^2}} dx = -\frac{\sqrt{49-36x^2}}{49x} + C.$

Solution.

a. By (1) in Appendix A with $a = 2$ and $u = x$,

$$\int \sqrt{x^2 + 4} dx = \frac{1}{2} \arctan\left(\frac{x}{2}\right) + C.$$

b. Let $u = x^2 + 4$, so $du = 2x dx$. Thus

$$\int \frac{x}{\sqrt{x^2+4}} dx = \frac{1}{2} \int \frac{du}{\sqrt{u}} = \frac{1}{2} \cdot 2u^{1/2} + C = \sqrt{x^2 + 4} + C.$$

c. Letting $a = 4$ and $u = 5x$, we see $du = 5dx$. Thus,

$$\int \frac{2}{\sqrt{16 + 25x^2}} dx = \frac{2}{5} \int \frac{du}{\sqrt{a^2 + u^2}}.$$

Now by (2) in Appendix A, it follows

$$\int \frac{2}{\sqrt{16 + 25x^2}} dx = \frac{2}{5} \ln |5x + \sqrt{16 + 25x^2}| + C.$$

d. Letting $a = 7$ and $u = 6x$, it follows $x = \frac{1}{6}u$ and $du = 6dx$, and therefore

$$\int \frac{1}{x^2\sqrt{49 - 36x^2}} dx = \frac{1}{6} \int \frac{1}{\frac{1}{36}u^2\sqrt{a^2 - u^2}} du.$$

Using (11) in Appendix A and the fact that $\frac{1}{6} \cdot 36 = 6$, we see

$$\int \frac{1}{x^2\sqrt{49 - 36x^2}} dx = -6 \cdot \frac{\sqrt{49 - 36x^2}}{49 \cdot 6x} + C.$$

5.7.3 Using Computer Algebra Systems

A computer algebra system (CAS) is a computer program that is capable of executing symbolic mathematics. For example, if we ask a CAS to solve the equation $ax^2 + bx + c = 0$ for the variable x , where a , b , and c are arbitrary constants, the program will return $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. Research to develop the first CAS dates to the 1960s, and these programs became publicly available in the early 1990s. Two prominent examples are the programs *Maple* and *Mathematica*, which were among the first computer algebra systems to offer a graphical user interface. Today, *Maple* and *Mathematica* are exceptionally powerful professional software packages that can execute an amazing array of sophisticated mathematical computations. They are also very expensive, as each is a proprietary program. The CAS *SAGE* is an open-source, free alternative to *Maple* and *Mathematica*.

For the purposes of this text, when we need to use a CAS, we are going to turn instead to a similar, but somewhat different computational tool, the web-based “computational knowledge engine” called *WolframAlpha*. There are two features of *WolframAlpha* that make it stand out from the CAS options mentioned above: (1) unlike *Maple* and *Mathematica*, *WolframAlpha* is free (provided we are willing to navigate some pop-up advertising); and (2) unlike any of the three, the syntax in *WolframAlpha* is flexible. Think of *WolframAlpha* as being a little bit like doing a Google search: the program will interpret what is input, and then provide a summary of options.

If we want to have *WolframAlpha* evaluate an integral for us, we can provide it syntax such as

integrate x^2 dx

to which the program responds with

$$\int x^2 dx = \frac{x^3}{3} + \text{constant.}$$

While there is much to be enthusiastic about regarding CAS programs such as *WolframAlpha*, there are several things we should be cautious about: (1) a CAS only responds to exactly what is input; (2) a CAS can answer using powerful functions from very advanced mathematics; and (3) there are problems that even a CAS cannot do without additional human insight.

Although (1) likely goes without saying, we have to be careful with our input: if we enter syntax that defines the wrong function, the CAS will work with precisely the function we define. For example, if we are interested in evaluating the integral

$$\int \frac{1}{16 - 5x^2} dx,$$

and we mistakenly enter

```
integrate 1/16 - 5x^2 dx
```

a CAS will (correctly) reply with

$$\frac{1}{16}x - \frac{5}{3}x^3.$$

But if we are sufficiently well-versed in antiderivatives, we will recognize that this function cannot be the one that we seek: integrating a rational function such as $\frac{1}{16-5x^2}$, we expect the logarithm function to be present in the result.

Regarding (2), even for a relatively simple integral such as $\int \frac{1}{16-5x^2} dx$, some CASs will invoke advanced functions rather than simple ones. For instance, if we use *Maple* to execute the command

```
int(1/(16-5*x^2), x);
```

the program responds with

$$\int \frac{1}{16 - 5x^2} dx = \frac{\sqrt{5}}{20} \operatorname{arctanh}\left(\frac{\sqrt{5}}{4}x\right).$$

While this is correct (save for the missing arbitrary constant, which *Maple* never reports), the inverse hyperbolic tangent function is not a common nor familiar one; a simpler way to express this function can be found by using the partial fractions method, and happens to be the result reported by *WolframAlpha*:

$$\int \frac{1}{16 - 5x^2} dx = \frac{1}{8\sqrt{5}} \left(\log(4\sqrt{5} + 5x) - \log(4\sqrt{5} - 5x) \right) + \text{constant.}$$

Using sophisticated functions from more advanced mathematics is sometimes the way a CAS says to the user “I don’t know how to do this problem.” For example, if we want to evaluate

$$\int e^{-x^2} dx,$$

and we ask *WolframAlpha* to do so, the input

```
integrate exp(-x^2) dx
```

results in the output

$$\int e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \operatorname{erf}(x) + \text{constant.}$$

The function “ $\operatorname{erf}(x)$ ” is the *error function*, which is actually defined by an integral:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

So, in producing output involving an integral, the CAS has basically reported back to us the very question we asked.

Finally, as remarked at (3) above, there are times that a CAS will actually fail without some additional human insight. If we consider the integral

$$\int (1+x)e^x \sqrt{1+x^2e^{2x}} dx$$

and ask *WolframAlpha* to evaluate

```
int (1+x) * exp(x) * sqrt(1+x^2 * exp(2x)) dx,
```

the program thinks for a moment and then reports

(no result found in terms of standard mathematical functions)

But in fact this integral is not that difficult to evaluate. If we let $u = xe^x$, then $du = (1+x)e^x dx$, which means that the preceding integral has form

$$\int (1+x)e^x \sqrt{1+x^2e^{2x}} dx = \int \sqrt{1+u^2} du,$$

which is a straightforward one for any CAS to evaluate.

So, we should proceed with some caution: while any CAS is capable of evaluating a wide range of integrals (both definite and indefinite), there are times when the result can mislead us. We must think carefully about the meaning of the output, whether it is consistent with what we expect, and whether or not it makes sense to proceed.

5.7.4 Summary

- We can antidifferentiate any rational function with the method of partial fractions. Any polynomial function can be factored into a product of linear and irreducible quadratic terms, so any rational function may be written as the sum of a polynomial plus rational terms of the form $\frac{A}{(x-c)^n}$ (where n is a natural number) and $\frac{Bx+C}{x^2+k}$ (where k is a positive real number).

- Until the development of computer algebra systems, integral tables enabled students of calculus to evaluate integrals such as $\int \sqrt{a^2 + u^2} du$, where a is a positive real number. A short table of integrals may be found in Appendix A.
- Computer algebra systems can play an important role in finding antiderivatives, though we must be cautious to use correct input, to watch for unusual or unfamiliar advanced functions that the CAS may cite in its result, and to consider the possibility that a CAS may need further assistance or insight from us in order to answer a particular question.

5.7.5 Exercises

- (Partial fractions: linear over difference of squares) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Partial fractions: constant over product) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Partial fractions: linear over quadratic) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Partial fractions: cubic over 4th degree) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Partial fractions: quadratic over factored cubic) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- For each of the following integrals involving rational functions, (1) use a CAS to find the partial fraction decomposition of the integrand; (2) evaluate the integral of the resulting function without the assistance of technology; (3) use a CAS to evaluate the original integral to test and compare your result in (2).

a. $\int \frac{x^3+x+1}{x^4-1} dx$

b. $\int \frac{x^5+x^2+3}{x^3-6x^2+11x-6} dx$

c. $\int \frac{x^2-x-1}{(x-3)^3} dx$

Answer. Exercise Answer

Solution.

a. Wolfram | Alpha shows that

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

Then, using standard antiderivatives,

$$\begin{aligned} \int \frac{x^3 + x + 1}{x^4 - 1} dx &= \int -\frac{1}{2(x^2 + 1)} + \frac{1}{4(x + 1)} + \frac{3}{4(x - 1)} dx \\ &= \int -\frac{1}{2(x^2 + 1)} dx + \int \frac{1}{4(x + 1)} dx + \text{int} \frac{3}{4(x - 1)} dx \\ &= -\frac{1}{2} \arctan(x) + \frac{1}{4} \ln|x + 1| + \frac{3}{4} \ln|x - 1| + C \end{aligned}$$

Directly asking Wolfram | Alpha to integrate $\frac{x^3+x+1}{x^4-1}$ yields

$$\int \frac{x^3+x+1}{x^4-1} dx = \frac{1}{4} (3 \log(1-x) + \log(x+1) - 2 \arctan x) + C.$$

By rearranging, distributing the $\frac{1}{4}$, and noticing that Wolfram | Alpha does not use absolute value with integrals involving the logarithm (as well as that $|x-1| = |1-x|$), we see that these two results are the same.

- b. Wolfram | Alpha shows that

$$\frac{x^5+x^2+3}{x^3-6x^2+11x-6} = x^2+6x+25 + \frac{255}{2(x-3)} - \frac{39}{x-2} + \frac{5}{2(x-1)}.$$

Then, using standard antiderivatives,

$$\begin{aligned} \int \frac{x^5+x^2+3}{x^3-6x^2+11x-6} dx &= \int x^2+6x+25 + \frac{255}{2(x-3)} - \frac{39}{x-2} + \frac{5}{2(x-1)} dx \\ &= \int x^2 dx + \int 6x dx + \int 25 dx + \int \frac{255}{2(x-3)} dx - \int \frac{39}{x-2} dx + \int \frac{5}{2(x-1)} dx \\ &= \frac{x^3}{3} + 3x^2 + 25x + \frac{255}{2} \ln|x-3| - 39 \ln|x-2| + \frac{5}{2} \ln|x-1| + C. \end{aligned}$$

Directly asking Wolfram | Alpha to integrate $\frac{x^5+x^2+3}{x^3-6x^2+11x-6}$ yields the same result with the caveats mentioned in part (a).

- c. Wolfram | Alpha shows that

$$\frac{x^2-x-1}{(x-3)^3} = \frac{1}{x-3} + \frac{5}{(x-3)^2} + \frac{5}{(x-3)^3}.$$

Then, using standard antiderivatives,

$$\begin{aligned} \int \frac{x^2-x-1}{(x-3)^3} dx &= \int \frac{1}{x-3} + \frac{5}{(x-3)^2} + \frac{5}{(x-3)^3} dx \\ &= \int \frac{1}{x-3} dx + \int \frac{5}{(x-3)^2} dx + \int \frac{5}{(x-3)^3} dx \\ &= \ln|x-3| + \int 5(x-3)^{-2} dx + \int 5(x-3)^{-3} dx \\ &= \ln|x-3| - \frac{5}{x-3} - \frac{5}{2(x-3)^2} + C. \end{aligned}$$

Directly asking Wolfram | Alpha to integrate $\frac{x^2-x-1}{(x-3)^3}$ yields

$$\int \frac{x^2-x-1}{(x-3)^3} dx = \log(x-3) - \frac{5(2x-5)}{2(x-3)^2} + C.$$

The logarithm part is the same, and since

$$-\frac{5}{x-3} - \frac{5}{2(x-3)^2} = -\frac{5[2(x-3)+1]}{2(x-3)^2} = -\frac{5(2x-5)}{2(x-3)^2},$$

we see that the rest of the result matches as well.

7. For each of the following integrals involving radical functions, (1) use an appropriate u -substitution along with Appendix A to evaluate the integral without the assistance of technology, and (2) use a CAS to evaluate the original integral to test and compare your result in (1).

- $\int \frac{1}{x\sqrt{9x^2+25}} dx$
- $\int x\sqrt{1+x^4} dx$
- $\int e^x\sqrt{4+e^{2x}} dx$
- $\int \frac{\tan(x)}{\sqrt{9-\cos^2(x)}} dx$

Answer.

- $\int \frac{1}{x\sqrt{9x^2+25}} dx = -\frac{1}{5} \ln \left| \frac{5+\sqrt{9x^2+5^2}}{3x} \right| + C.$
- $\int x\sqrt{1+x^4} dx = \frac{1}{2} \left(\frac{x^2}{2} \sqrt{x^4+1} + \frac{1}{2} \ln |x^2 + \sqrt{x^4+1}| \right) + C$
- $\int e^x\sqrt{4+e^{2x}} dx = \frac{e^x}{2} \sqrt{e^{2x}+4} + 2 \ln |e^x + \sqrt{e^{2x}+4}| + C$
- $\int \frac{\tan(x)}{\sqrt{9-\cos^2(x)}} dx = \frac{1}{3} \ln \left| \frac{3+\sqrt{9-\cos^2(x)}}{\cos(x)} \right| + C.$

Solution.

- For $\int \frac{1}{x\sqrt{9x^2+25}} dx$, we let $u = 3x$ so $du = 3 dx$ and $x = \frac{1}{3}u$. Applying this substitution,

$$\int \frac{1}{x\sqrt{9x^2+25}} dx = \int \frac{3}{u\sqrt{u^2+25}} \cdot \frac{1}{3} du = \int \frac{1}{u\sqrt{u^2+25}} du.$$

Now, using Rule (e) from the appendix with $a = 5$,

$$\int \frac{1}{u\sqrt{u^2+25}} du = -\frac{1}{5} \ln \left| \frac{5+\sqrt{u^2+5^2}}{u} \right| + C$$

Substituting back to x , we've shown that

$$\int \frac{1}{x\sqrt{9x^2+25}} dx = -\frac{1}{5} \ln \left| \frac{5+\sqrt{9x^2+5^2}}{3x} \right| + C.$$

Asking Wolfram|Alpha to evaluate the same original integral, the program reports that

$$\int \frac{1}{x\sqrt{9x^2+25}} dx = \frac{1}{5} (\log(x) - \log(\sqrt{9x^2+25} + 5)) + C.$$

These two results are in fact the same (the program uses “log” where we use “ln”), though it takes some work with logarithm properties to show this. Using the rule $\ln(\frac{u}{v}) = \ln(u) - \ln(v)$ and $\ln(3x) = \ln(3) + \ln(x)$ and including the value $\ln(3)$ in the arbitrary constant of integration, it follows that the two results match.

- For $\int x\sqrt{1+x^4} dx$, we first recognize that $x^4 = (x^2)^2$, and that with $u = x^2$ we have a function-derivative pair present with $du = 2x dx$. Applying this substitution,

$$\int x\sqrt{1+x^4} dx = \int \frac{1}{2} \sqrt{1+u^2} du.$$

Next, we use Rule (c) with “+” and $a = 1$ to find that

$$\int \frac{1}{2} \sqrt{1+u^2} du = \frac{1}{2} \left(\frac{u}{2} \sqrt{u^2+1} + \frac{1}{2} \ln |u + \sqrt{u^2+1}| \right) + C$$

Returning to the original variable, x , we have

$$\int x \sqrt{1+x^4} dx = \frac{1}{2} \left(\frac{x^2}{2} \sqrt{x^4+1} + \frac{1}{2} \ln |x^2 + \sqrt{x^4+1}| \right) + C$$

Wolfram | Alpha's first response when asked to evaluate this integral is to give a result that involves the hyperbolic sine function ($\sinh^{-1}(x^2)$); if we scroll down to “alternate results”, we find one that matches what we've found above.

- c. For $\int e^x \sqrt{4+e^{2x}} dx$, we first observe that $e^{2x} = (e^x)^2$. If we use the substitution $u = e^x$, the resulting integral is

$$\int e^x \sqrt{4+e^{2x}} dx = \int \sqrt{4+u^2} du,$$

which can be evaluated by the same approach as in part (b) with Rule (c) from the Appendix, this time with $a = 2$. Thus,

$$\int \sqrt{4+u^2} du = \frac{u}{2} \sqrt{u^2+4} + \frac{2^2}{2} \ln |u + \sqrt{u^2+4}|$$

Substituting back,

$$\int e^x \sqrt{4+e^{2x}} dx = \frac{e^x}{2} \sqrt{e^{2x}+4} + 2 \ln |e^x + \sqrt{e^{2x}+4}| + C$$

As in part (b), we need to scroll down in Wolfram | Alpha's results to find one close to what we have here. The alternate result almost matches, but the program writes

$$\log\left(\sqrt{\frac{e^{2x}}{4}+1} + \frac{e^x}{2}\right)$$

where we have

$$\ln |e^x + \sqrt{e^{2x}+4}|$$

This discrepancy can be explained using properties of the logarithm function and the presence of the arbitrary constant. If we write

$$\log\left(\sqrt{\frac{e^{2x}}{4}+1} + \frac{e^x}{2}\right) = \log\left(\sqrt{\frac{e^{2x}+4}{4}} + \frac{e^x}{2}\right) = \log\left(\frac{\sqrt{e^{2x}+4}+e^x}{2}\right)$$

and then recall that $\log\left(\frac{z}{2}\right) = \log(z) - \log(2)$, we see that the $\log(2)$ can be included in the arbitrary constant, and the two different results actually match.

- d. To evaluate the integral $\int \frac{\tan(x)}{\sqrt{9-\cos^2(x)}} dx$, we first write $\tan(x) = \frac{\sin(x)}{\cos(x)}$ to have the integral in the form

$$\int \frac{\tan(x)}{\sqrt{9-\cos^2(x)}} dx = \int \frac{\sin(x)}{\cos(x)\sqrt{9-\cos^2(x)}} dx.$$

Next, letting $u = \cos(x)$, it follows $du = -\sin(x) dx$, and therefore

$$\int \frac{\tan(x)}{\sqrt{9 - \cos^2(x)}} dx = \int \frac{-1}{u\sqrt{9 - u^2}} du.$$

By Rule (j) in the appendix, we know that

$$\int \frac{1}{u\sqrt{9 - u^2}} du = -\frac{1}{3} \ln \left| \frac{3 + \sqrt{9 - u^2}}{u} \right| + C.$$

Connecting the above chain of equalities and substituting back to x , we have shown that

$$\int \frac{\tan(x)}{\sqrt{9 - \cos^2(x)}} dx = \frac{1}{3} \ln \left| \frac{3 + \sqrt{9 - \cos^2(x)}}{\cos(x)} \right| + C.$$

When Wolfram | Alpha is asked to evaluate the original integral, its first result involves the inverse hyperbolic tangent function, and the alternate forms are equally complicated. It seems that the technology is unable to evaluate this integral in the most straightforward way.

8. Consider the indefinite integral given by

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

- Explain why u -substitution does not offer a way to simplify this integral by discussing at least two different options you might try for u .
- Explain why integration by parts does not seem to be a reasonable way to proceed, either, by considering one option for u and dv .
- Is there any line in the integral table in Appendix A that is helpful for this integral?
- Evaluate the given integral using *WolframAlpha*. What do you observe?

Answer.

- Try $u = 1 + x^2$ or $u = x + \sqrt{1 + x^2}$.
- Try $u = \sqrt{x + \sqrt{1 + x^2}}$ and $dv = \frac{1}{x} dx$.
- No.
- It appears that the function $\frac{\sqrt{x + \sqrt{1 + x^2}}}{x}$ does not have an elementary antiderivative.

Solution.

- Looking at the composite functions that are present, there are two possible u -substitutions to try. One is $u = 1 + x^2$, for which $du = 2x dx$ and $x = \sqrt{u - 1}$. Because there is no $x dx$ term present in the integral, this substitution will not work.

A second substitution to attempt is letting $u = x + \sqrt{1 + x^2}$, for which $du = (1 + x(1 + x^2)^{-1/2}) dx$. Here, too, the needed product is not present in the integral in order for the substitution to work.

- b. For integration by parts to work, we normally need a choice of u that is straightforward to differentiate, and dv that is straightforward to antidifferentiate. Noting that the integrand can really only be written as the product of $\frac{1}{x}$ and $\sqrt{x + \sqrt{1 + x^2}}$, it appears that the most natural possible option is to let $u = \sqrt{x + \sqrt{1 + x^2}}$ and $dv = \frac{1}{x} dx$, since we can antidifferentiate $\frac{1}{x}$. Here, however, the approach stalls, because integration by parts requires us to evaluate $\int v du$, and that integral is

$$\int \ln|x| \cdot \frac{1}{2} (x + \sqrt{1 + x^2})^{-1/2} \cdot (1 + x(1 + x^2)^{-1/2}) dx$$

which is far more complicated than the already-complicated integral we started with. From this it appears that integration by parts will not be productive.

- c. Since we can't seem to find a u -substitution that simplifies the given integral, and there are no formulas in the Appendix that use nested square roots, it appears that the table doesn't offer any help in evaluating this particular integral.
- d. Wolfram | Alpha returns results that either involve inverse hyperbolic trigonometric functions or complex numbers. As such, it appears that the function $\frac{\sqrt{x+\sqrt{1+x^2}}}{x}$ does not have an elementary antiderivative that we can find.

5.8 Numerical Integration

Motivating Questions

- How do we accurately evaluate a definite integral such as $\int_0^1 e^{-x^2} dx$ when we cannot use the First Fundamental Theorem of Calculus because the integrand lacks an elementary algebraic antiderivative? Are there ways to generate accurate estimates without using extremely large values of n in Riemann sums?
 - What is the Trapezoid Rule, and how is it related to left, right, and middle Riemann sums?
 - How are the errors in the Trapezoid Rule and Midpoint Rule related, and how can they be used to develop an even more accurate rule?
-

When we first explored finding the net signed area bounded by a curve, we developed the concept of a Riemann sum as a helpful estimation tool and a key step in the definition of the definite integral. Recall that the left, right, and middle Riemann sums of a function f on an interval $[a, b]$ are given by

$$L_n = f(x_0)\Delta x + f(x_1)\Delta x + \cdots + f(x_{n-1})\Delta x = \sum_{i=0}^{n-1} f(x_i)\Delta x, \quad (5.8.1)$$

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = \sum_{i=1}^n f(x_i)\Delta x, \quad (5.8.2)$$

$$M_n = f(\bar{x}_1)\Delta x + f(\bar{x}_2)\Delta x + \cdots + f(\bar{x}_n)\Delta x = \sum_{i=1}^n f(\bar{x}_i)\Delta x, \quad (5.8.3)$$

where $x_0 = a$, $x_i = a + i\Delta x$, $x_n = b$, and $\Delta x = \frac{b-a}{n}$. For the middle sum, we defined $\bar{x}_i = (x_{i-1} + x_i)/2$.

A Riemann sum is a sum of (possibly signed) areas of rectangles. The value of n determines the number of rectangles, and our choice of left endpoints, right endpoints, or midpoints determines the heights of the rectangles. We can see the similarities and differences among these three options in Figure 5.8.1, where we consider the function $f(x) = \frac{1}{20}(x-4)^3 + 7$ on the interval $[1, 8]$, and use 5 rectangles for each of the Riemann sums.

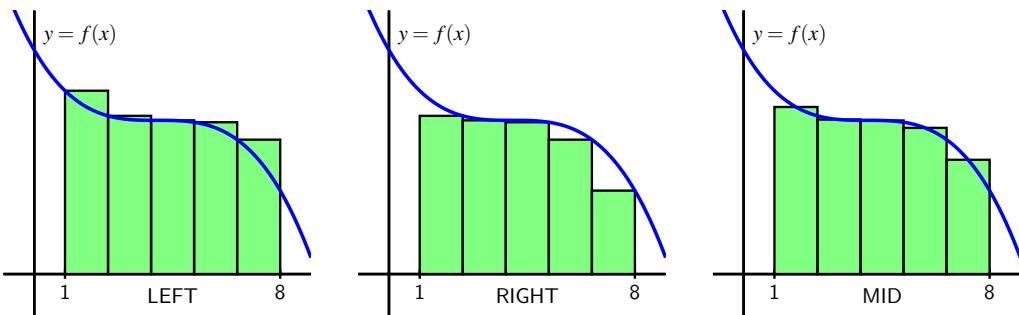


Figure 5.8.1: Left, right, and middle Riemann sums for $y = f(x)$ on $[1, 8]$ with 5 subintervals.

While it is a good exercise to compute a few Riemann sums by hand, just to ensure that we understand how they work and how varying the function, the number of subintervals, and the choice of endpoints or midpoints affects the result, using computing technology is the best way to determine L_n , R_n , and M_n . Any computer algebra system will offer this capability; as we saw in Example 4.3.2, a straightforward option that is freely available online is the applet¹ at <http://gvsu.edu/s/a9>. Note that we can adjust the formula for $f(x)$, the window of x - and y -values of interest, the number of subintervals, and the method. (See Example 4.3.2 for any needed reminders on how the applet works.)

In this section we explore several different alternatives for estimating definite integrals. Our main goal is to develop formulas to estimate definite integrals accurately without using a large numbers of rectangles.

Example 5.8.2. As we begin to investigate ways to approximate definite integrals, it will be insightful to compare results to integrals whose exact values we know. To that end, the following sequence of questions centers on $\int_0^3 x^2 dx$.

- Use the applet at <http://gvsu.edu/s/a9> with the function $f(x) = x^2$ on the window of x values from 0 to 3 to compute L_3 , the left Riemann sum with three subintervals.
- Likewise, use the applet to compute R_3 and M_3 , the right and middle Riemann sums with three subintervals, respectively.
- Use the Fundamental Theorem of Calculus to compute the exact value of $I = \int_0^3 x^2 dx$.
- We define the *error* in an approximation of a definite integral to be the difference between the integral's exact value and the approximation's value. What is the error that results from using L_3 ? From R_3 ? From M_3 ?
- In what follows in this section, we will learn a new approach to estimating the value of a definite integral known as the Trapezoid Rule. The basic idea is to use trapezoids, rather than rectangles, to estimate the area under a curve. What is the formula for the area of a trapezoid with bases of length b_1 and b_2 and height h ?

¹Marc Renault, Shippensburg University

- f. Working by hand, estimate the area under $f(x) = x^2$ on $[0, 3]$ using three subintervals and three corresponding trapezoids. What is the error in this approximation? How does it compare to the errors you calculated in (d)?

5.8.1 The Trapezoid Rule

So far, we have used the simplest possible quadrilaterals (that is, rectangles) to estimate areas. It is natural, however, to wonder if other familiar shapes might serve us even better.

An alternative to L_n , R_n , and M_n is called the *Trapezoid Rule*. Rather than using a rectangle to estimate the (signed) area bounded by $y = f(x)$ on a small interval, we use a trapezoid. For example, in Figure 5.8.3, we estimate the area under the curve using three subintervals and the trapezoids that result from connecting the corresponding points on the curve with straight lines.

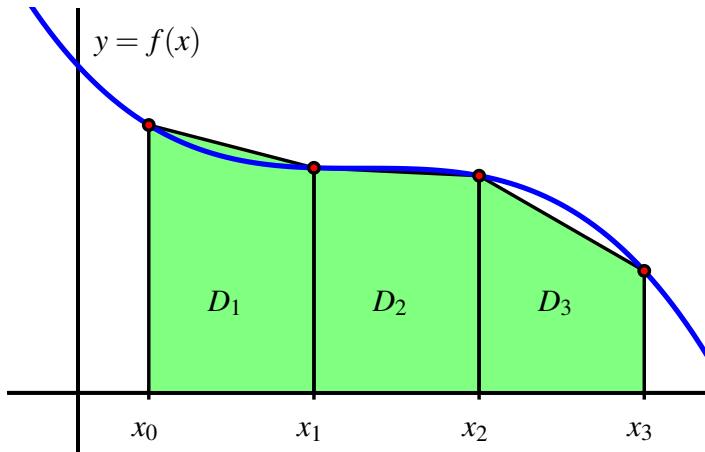


Figure 5.8.3: Estimating $\int_a^b f(x) dx$ using three subintervals and trapezoids, rather than rectangles, where $a = x_0$ and $b = x_3$.

The biggest difference between the Trapezoid Rule and a Riemann sum is that on each subinterval, the Trapezoid Rule uses two function values, rather than one, to estimate the (signed) area bounded by the curve. For instance, to compute D_1 , the area of the trapezoid on $[x_0, x_1]$, we observe that the left base has length $f(x_0)$, while the right base has length $f(x_1)$. The height of the trapezoid is $x_1 - x_0 = \Delta x = \frac{b-a}{3}$. The area of a trapezoid is the average of the bases times the height, so we have

$$D_1 = \frac{1}{2}(f(x_0) + f(x_1)) \cdot \Delta x.$$

Using similar computations for D_2 and D_3 , we find that T_3 , the trapezoidal approximation

to $\int_a^b f(x) dx$ is given by

$$\begin{aligned} T_3 &= D_1 + D_2 + D_3 \\ &= \frac{1}{2}(f(x_0) + f(x_1)) \cdot \Delta x + \frac{1}{2}(f(x_1) + f(x_2)) \cdot \Delta x + \frac{1}{2}(f(x_2) + f(x_3)) \cdot \Delta x. \end{aligned}$$

Because both left and right endpoints are being used, we recognize within the trapezoidal approximation the use of both left and right Riemann sums. Rearranging the expression for T_3 by removing factors of $\frac{1}{2}$ and Δx , grouping the left endpoint and right endpoint evaluations of f , we see that

$$T_3 = \frac{1}{2} [f(x_0) + f(x_1) + f(x_2)] \Delta x + \frac{1}{2} [f(x_1) + f(x_2) + f(x_3)] \Delta x. \quad (5.8.4)$$

We now observe that two familiar sums have arisen. The left Riemann sum L_3 is $L_3 = f(x_0)\Delta x + f(x_1)\Delta x + f(x_2)\Delta x$, and the right Riemann sum is $R_3 = f(x_1)\Delta x + f(x_2)\Delta x + f(x_3)\Delta x$. Substituting L_3 and R_3 for the corresponding expressions in Equation (5.8.4), it follows that $T_3 = \frac{1}{2} [L_3 + R_3]$. We have thus seen a very important result: using trapezoids to estimate the (signed) area bounded by a curve is the same as averaging the estimates generated by using left and right endpoints.

The Trapezoid Rule

The trapezoidal approximation, T_n , of the definite integral $\int_a^b f(x) dx$ using n subintervals is given by the rule

$$\begin{aligned} T_n &= \left[\frac{1}{2}(f(x_0) + f(x_1)) + \frac{1}{2}(f(x_1) + f(x_2)) + \cdots + \frac{1}{2}(f(x_{n-1}) + f(x_n)) \right] \Delta x. \\ &= \sum_{i=0}^{n-1} \frac{1}{2}(f(x_i) + f(x_{i+1})) \Delta x. \end{aligned}$$

Moreover, $T_n = \frac{1}{2} [L_n + R_n]$.

Example 5.8.4. In this example, we explore the relationships among the errors generated by left, right, midpoint, and trapezoid approximations to the definite integral $\int_1^2 \frac{1}{x^2} dx$

- Use the First FTC to evaluate $\int_1^2 \frac{1}{x^2} dx$ exactly.
- Use appropriate computing technology to compute the following approximations for $\int_1^2 \frac{1}{x^2} dx$: T_4 , M_4 , T_8 , and M_8 .
- Let the *error* of an approximation be the difference between the exact value of the definite integral and the resulting approximation. For instance, if we let $E_{T,4}$ represent the error that results from using the trapezoid rule with 4 subintervals to estimate the integral, we have

$$E_{T,4} = \int_1^2 \frac{1}{x^2} dx - T_4.$$

Similarly, we compute the error of the midpoint rule approximation with 8 subintervals by the formula

$$E_{M,8} = \int_1^2 \frac{1}{x^2} dx - M_8.$$

Based on your work in (a) and (b) above, compute $E_{T,4}$, $E_{T,8}$, $E_{M,4}$, $E_{M,8}$.

- d. Which rule consistently over-estimates the exact value of the definite integral? Which rule consistently under-estimates the definite integral?
- e. What behavior(s) of the function $f(x) = \frac{1}{x^2}$ lead to your observations in (d)?

Hint.

- a. $\frac{1}{x^2} = x^{-2}$.
- b. Use a computational device.
- c. Use a computational device.
- d. Which estimate is larger than the true value of the definite integral?
- e. Note that how the curve bends makes a big difference in whether the trapezoid rule over- or under-estimates the value of the definite integral.

Answer.

- a. $\int_1^2 \frac{1}{x^2} dx = \frac{1}{2}$.
- b. The table below gives values of the trapezoid rule and corresponding errors for different n -values.

n	T_n	$E_{T,n}$
4	0.50899	-0.50899
8	0.50227	-0.50227
16	0.50057	-0.50057

- c. The table below gives values of the midpoint rule and corresponding errors for different n -values.

n	M_n	$E_{M,n}$
4	0.49555	0.00445
8	0.49887	0.00113
16	0.49972	0.00028

- d. The trapezoid rule overestimates; the midpoint rule underestimates.
- e. $f(x) = \frac{1}{x^2}$ is concave up on $[1, 2]$.

Solution.

a. $\int_1^2 \frac{1}{x^2} dx = -x^{-1}|_1^2 = -\frac{1}{2} + 1 = \frac{1}{2}.$

- b. The table below gives values of the trapezoid rule and corresponding errors for different n -values.

n	T_n	$E_{T,n}$
4	0.50899	-0.50899
8	0.50227	-0.50227
16	0.50057	-0.50057

- c. The table below gives values of the midpoint rule and corresponding errors for different n -values.

n	M_n	$E_{M,n}$
4	0.49555	0.00445
8	0.49887	0.00113
16	0.49972	0.00028

- d. From the errors in comparison to the known exact value, we see that the trapezoid rule overestimates this definite integral and the midpoint rule underestimates this definite integral.
- e. The graph of the function given by $f(x) = \frac{1}{x^2}$ is concave up on the interval $[1, 2]$. Because of this fact, we can see graphically that the line forming the top of each trapezoid lies fully above the curve, and thus the trapezoid rule overestimates the true value of the definite integral. Later in this section we'll see graphically why this concavity makes the midpoint rule an underestimate.

5.8.2 Comparing the Midpoint and Trapezoid Rules

We know from the definition of the definite integral that if we let n be large enough, we can make any of the approximations L_n , R_n , and M_n as close as we'd like (in theory) to the exact value of $\int_a^b f(x) dx$. Thus, it may be natural to wonder why we ever use any rule other than L_n or R_n (with a sufficiently large n value) to estimate a definite integral. One of the primary reasons is that as $n \rightarrow \infty$, $\Delta x = \frac{b-a}{n} \rightarrow 0$, and thus in a Riemann sum calculation with a large n value, we end up multiplying by a number that is very close to zero. Doing so often generates roundoff error, because representing numbers close to zero accurately is a persistent challenge for computers.

Hence, we explore ways to estimate definite integrals to high levels of precision, but without using extremely large values of n . Paying close attention to patterns in errors, such as those observed in Example 5.8.4, is one way to begin to see some alternate approaches.

To begin, we compare the errors in the Midpoint and Trapezoid rules. First, consider a function that is concave up on a given interval, and picture approximating the area bounded on that interval by both the Midpoint and Trapezoid rules using a single subinterval.

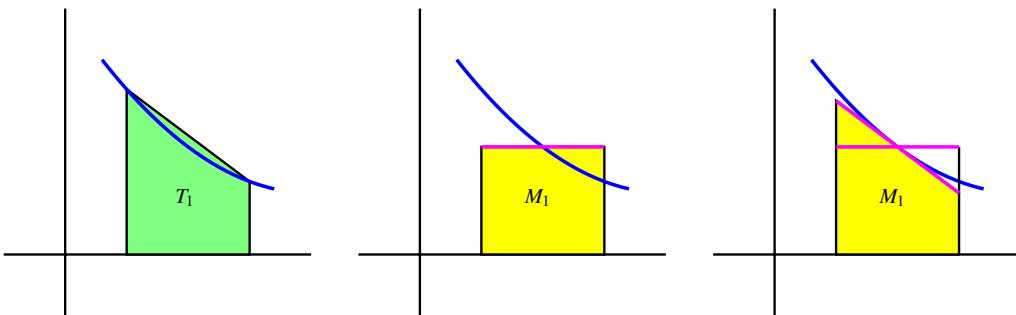


Figure 5.8.5: Estimating $\int_a^b f(x) dx$ using a single subinterval: at left, the trapezoid rule; in the middle, the midpoint rule; at right, a modified way to think about the midpoint rule.

As seen in Figure 5.8.5, it is evident that whenever the function is concave up on an interval, the Trapezoid Rule with one subinterval, T_1 , will overestimate the exact value of the definite integral on that interval. From a careful analysis of the line that bounds the top of the rectangle for the Midpoint Rule (shown in magenta), we see that if we rotate this line segment until it is tangent to the curve at the midpoint of the interval (as shown at right in Figure 5.8.5), the resulting trapezoid has the same area as M_1 , and this value is less than the exact value of the definite integral. Thus, when the function is concave up on the interval, M_1 underestimates the integral's true value.

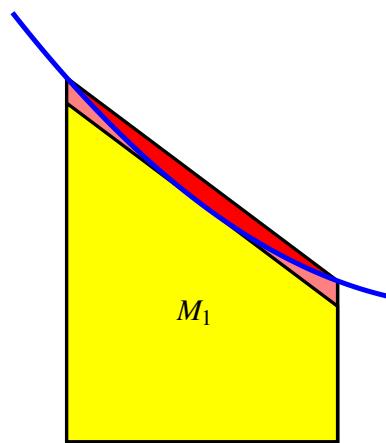


Figure 5.8.6: Comparing the error in estimating $\int_a^b f(x) dx$ using a single subinterval: in red, the error from the Trapezoid rule; in light red, the error from the Midpoint rule.

These observations extend easily to the situation where the function's concavity remains consistent but we use larger values of n in the Midpoint and Trapezoid Rules. Hence, whenever f is concave up on $[a, b]$, T_n will overestimate the value of $\int_a^b f(x) dx$, while M_n will

underestimate $\int_a^b f(x) dx$. The reverse observations are true in the situation where f is concave down.

Next, we compare the size of the errors between M_n and T_n . Again, we focus on M_1 and T_1 on an interval where the concavity of f is consistent. In Figure 5.8.6, where the error of the Trapezoid Rule is shaded in red, while the error of the Midpoint Rule is shaded lighter red, it is visually apparent that the error in the Trapezoid Rule is more significant. To see how much more significant, let's consider two examples and some particular computations.

If we let $f(x) = 1 - x^2$ and consider $\int_0^1 f(x) dx$, we know by the First FTC that the exact value of the integral is

$$\int_0^1 (1 - x^2) dx = x - \frac{x^3}{3} \Big|_0^1 = \frac{2}{3}.$$

Using appropriate technology to compute M_4 , M_8 , T_4 , and T_8 , as well as the corresponding errors $E_{M,4}$, $E_{M,8}$, $E_{T,4}$, and $E_{T,8}$, as we did in Example 5.8.4, we find the results summarized in Table 5.8.7. We also include the approximations and their errors for the example $\int_1^2 \frac{1}{x^2} dx$ from Example 5.8.4.

Rule	$\int_0^1 (1 - x^2) dx = 0.6$	error	$\int_1^2 \frac{1}{x^2} dx = 0.5$	error
T_4	0.65625	-0.0104166667	0.5089937642	0.0089937642
M_4	0.671875	0.0052083333	0.4955479365	-0.0044520635
T_8	0.6640625	-0.0026041667	0.5022708502	0.0022708502
M_8	0.66796875	0.0013020833	0.4988674899	-0.0011325101

Table 5.8.7: Calculations of T_4 , M_4 , T_8 , and M_8 , along with corresponding errors, for the definite integrals $\int_0^1 (1 - x^2) dx$ and $\int_1^2 \frac{1}{x^2} dx$.

For a given function f and interval $[a, b]$, $E_{T,4} = \int_a^b f(x) dx - T_4$ calculates the difference between the exact value of the definite integral and the approximation generated by the Trapezoid Rule with $n = 4$. If we look at not only $E_{T,4}$, but also the other errors generated by using T_n and M_n with $n = 4$ and $n = 8$ in the two examples noted in Table 5.8.7, we see an evident pattern. Not only is the sign of the error (which measures whether the rule generates an over- or under-estimate) tied to the rule used and the function's concavity, but the magnitude of the errors generated by T_n and M_n seems closely connected. In particular, the errors generated by the Midpoint Rule seem to be about half the size of those generated by the Trapezoid Rule.

That is, we can observe in both examples that $E_{M,4} \approx -\frac{1}{2}E_{T,4}$ and $E_{M,8} \approx -\frac{1}{2}E_{T,8}$. This property of the Midpoint and Trapezoid Rules turns out to hold in general: for a function of consistent concavity, the error in the Midpoint Rule has the opposite sign and approximately half the magnitude of the error of the Trapezoid Rule. Written symbolically,

$$E_{M,n} \approx -\frac{1}{2}E_{T,n}.$$

This important relationship suggests a way to combine the Midpoint and Trapezoid Rules to create an even more accurate approximation to a definite integral.

5.8.3 Simpson's Rule

When we first developed the Trapezoid Rule, we observed that it is an average of the Left and Right Riemann sums:

$$T_n = \frac{1}{2}(L_n + R_n).$$

If a function is always increasing or always decreasing on the interval $[a, b]$, one of L_n and R_n will over-estimate the true value of $\int_a^b f(x) dx$, while the other will under-estimate the integral. Thus, the errors found in L_n and R_n will have opposite signs; so averaging L_n and R_n eliminates a considerable amount of the error present in the respective approximations. In a similar way, it makes sense to think about averaging M_n and T_n in order to generate a still more accurate approximation.

We've just observed that M_n is typically about twice as accurate as T_n . So we use the weighted average

$$S_{2n} = \frac{2M_n + T_n}{3}. \quad (5.8.5)$$

The rule for S_{2n} giving by Equation (5.8.5) is usually known as *Simpson's Rule*.² Note that we use " S_{2n} " rather than " S_n " since the n points the Midpoint Rule uses are different from the n points the Trapezoid Rule uses, and thus Simpson's Rule is using $2n$ points at which to evaluate the function. We build upon the results in Table 5.8.7 to see the approximations generated by Simpson's Rule. In particular, in Table 5.8.8, we include all of the results in Table 5.8.7, but include additional results for $S_8 = \frac{2M_4+T_4}{3}$ and $S_{16} = \frac{2M_8+T_8}{3}$.

Rule	$\int_0^1 (1 - x^2) dx = 0.6$	error	$\int_1^2 \frac{1}{x^2} dx = 0.5$	error
T_4	0.65625	-0.0104166667	0.5089937642	0.0089937642
M_4	0.671875	0.0052083333	0.4955479365	-0.0044520635
S_8	0.6666666667	0	0.5000298792	0.0000298792
T_8	0.6640625	-0.0026041667	0.5022708502	0.0022708502
M_8	0.66796875	0.0013020833	0.4988674899	-0.0011325101
S_{16}	0.6666666667	0	0.5000019434	0.0000019434

Table 5.8.8: Table 5.8.7 updated to include S_8 , S_{16} , and the corresponding errors.

The results seen in Table 5.8.8 are striking. If we consider the S_{16} approximation of $\int_1^2 \frac{1}{x^2} dx$, the error is only $E_{S,16} = 0.0000019434$. By contrast, $L_8 = 0.5491458502$, so the error of that estimate is $E_{L,8} = -0.0491458502$. Moreover, we observe that generating the approximations for Simpson's Rule is almost no additional work: once we have L_n , R_n , and M_n for a given value of n , it is a simple exercise to generate T_n , and from there to calculate S_{2n} . Finally, note

²Thomas Simpson was an 18th century mathematician; his idea was to extend the Trapezoid rule, but rather than using straight lines to build trapezoids, to use quadratic functions to build regions whose area was bounded by parabolas (whose areas he could find exactly). Simpson's Rule is often developed from the more sophisticated perspective of using interpolation by quadratic functions.

that the error in the Simpson's Rule approximations of $\int_0^1 (1 - x^2) dx$ is zero!³

These rules are not only useful for approximating definite integrals such as $\int_0^1 e^{-x^2} dx$, for which we cannot find an elementary antiderivative of e^{-x^2} , but also for approximating definite integrals when we are given a function through a table of data.

Example 5.8.9. A car traveling along a straight road is braking and its velocity is measured at several different points in time, as given in the following table. Assume that v is continuous, always decreasing, and always decreasing at a decreasing rate, as is suggested by the data.

seconds, t	Velocity in ft/sec, $v(t)$
0	100
0.3	99
0.6	96
0.9	90
1.2	80
1.5	50
1.8	0

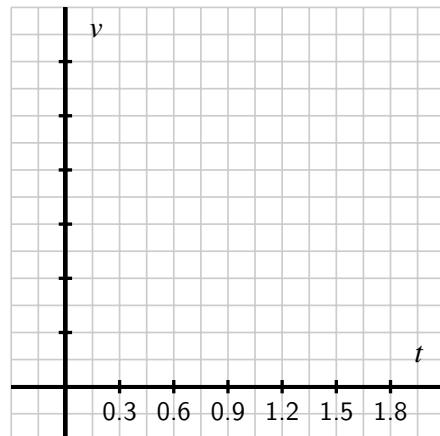


Table 5.8.10: Data for the braking car.

Figure 5.8.11: Axes for plotting the data in Example 5.8.9.

- Plot the given data on the set of axes provided in Figure 5.8.11 with time on the horizontal axis and the velocity on the vertical axis.
- What definite integral will give you the exact distance the car traveled on $[0, 1.8]$?
- Estimate the total distance traveled on $[0, 1.8]$ by computing L_3 , R_3 , and T_3 . Which of these under-estimates the true distance traveled?
- Estimate the total distance traveled on $[0, 1.8]$ by computing M_3 . Is this an over- or under-estimate? Why?
- Using your results from (c) and (d), improve your estimate further by using Simpson's Rule.

³Similar to how the Midpoint and Trapezoid approximations are exact for linear functions, Simpson's Rule approximations are exact for quadratic and cubic functions. See additional discussion on this issue later in the section and in the exercises.

- f. What is your best estimate of the average velocity of the car on $[0, 1.8]$? Why? What are the units on this quantity?

Hint.

- Plot the data.
- What are the units of $v(t) \cdot \Delta t$?
- Recall the standard rules for sums that produce L_3, R_3, T_3 .
- Think about concavity to decide if M_3 is an over- or under-estimate.
- Recall how S_3 is a weighted average of T_3 and M_3 .
- Simpson's Rule gives the best estimate for a function of consistent concavity.

Answer.

- a. Plot the data.

b. $\int_0^{1.8} v(t)dt$.

c.

$$L_3 = 165.6 \text{ ft}$$

$$R_3 = 105.6 \text{ ft}$$

$$T_3 = 135.6 \text{ ft}.$$

R_3 and T_3 are underestimates.

- d. $M_3 = 143.4 \text{ ft}$; overestimate.

e. $S_6 = 140.8 \text{ ft}$.

- f. Simpson's rule gives the best approximation of the distance traveled, $\int_0^{1.8} v(t)dt \approx 140.8 \text{ ft}$.

Solution.

- a. Plot the data.

- b. Since the velocity is always positive, the definite integral that will give the exact distance traveled by the car on the interval $[0, 1.8]$ is

$$\int_0^{1.8} v(t)dt.$$

- c. The estimates of $\int_0^{1.8} v(t)dt$ are

$$L_3 = 165.6 \text{ ft}$$

$$R_3 = 105.6 \text{ ft}$$

$$T_3 = 135.6 \text{ ft}.$$

R_3 is an underestimate of the distance traveled since $v(t)$ is decreasing. T_3 is an underestimate of the distance traveled since $v(t)$ is concave down.

- d. Another estimate of the distance traveled is

$$M_3 = 143.4 \text{ ft}.$$

This is an overestimate since $v(t)$ is concave down.

- e. For Simpson's Rule, we see that

$$S_6 = \frac{2}{3}M_3 + \frac{1}{3}T_3 = 140.8 \text{ ft}.$$

- f. Simpson's rule gives the best approximation of the distance traveled since it is a weighted average of the midpoint and trapezoid rules and uses more information about the velocity than the other methods. The units on each of the estimates, including Simpson's Rule, are "feet", since $\text{ft/sec} \cdot \text{sec} = \text{ft}$. Thus, the best approximation we have generated is that $\int_0^{1.8} v(t)dt \approx 140.8 \text{ ft}$.

5.8.4 Overall observations regarding L_n, R_n, T_n, M_n , and S_{2n} .

As we conclude our discussion of numerical approximation of definite integrals, it is important to summarize general trends in how the various rules over- or under-estimate the true value of a definite integral, and by how much. To revisit some past observations and see some new ones, we consider the following example.

Example 5.8.12. Consider the functions $f(x) = 2 - x^2$, $g(x) = 2 - x^3$, and $h(x) = 2 - x^4$, all on the interval $[0, 1]$. For each of the questions that require a numerical answer in what follows, write your answer exactly in fraction form.

- On the three sets of axes provided in Figure 5.8.13, sketch a graph of each function on the interval $[0, 1]$, and compute L_1 and R_1 for each. What do you observe?
- Compute M_1 for each function to approximate $\int_0^1 f(x)dx$, $\int_0^1 g(x)dx$, and $\int_0^1 h(x)dx$, respectively.
- Compute T_1 for each of the three functions, and hence compute S_2 for each of the three functions.
- Evaluate each of the integrals $\int_0^1 f(x)dx$, $\int_0^1 g(x)dx$, and $\int_0^1 h(x)dx$ exactly using the First FTC.
- For each of the three functions f , g , and h , compare the results of L_1 , R_1 , M_1 , T_1 , and S_2 to the true value of the corresponding definite integral. What patterns do you observe?

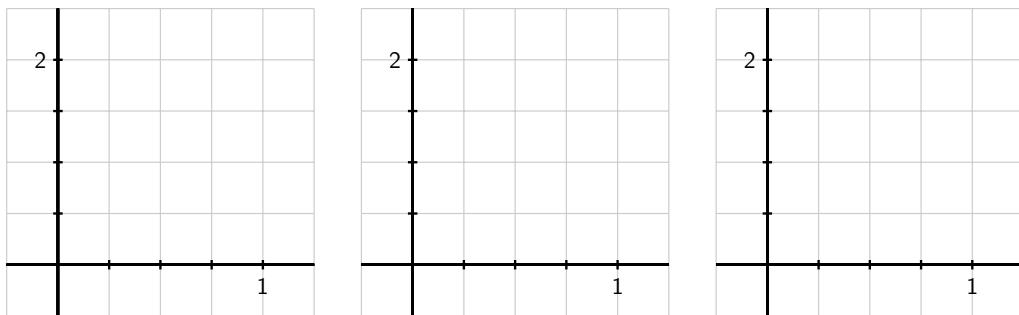


Figure 5.8.13: Axes for plotting the functions in Example 5.8.12.

Hint.

- For each estimate, just one function evaluation is needed.
- Use the midpoint rule with $n = 1$.
- Remember that both the trapezoid and Simpson's rule can be executed using (weighted) averages of known values.
- Find antiderivatives to evaluate the integrals exactly.
- Think about trends in over- and under-estimates.

Answer.

- For L_1 and T_1 :

f	g	h
$L_1 = 2$	$L_1 = 2$	$L_1 = 2$
$R_1 = 1$	$R_1 = 1$	$R_1 = 1$

Table 5.8.14: Left and Trapezoid rules.

The values of L_1 and R_1 are the same for all three.

- For the M_1 ,

f	g	h
$M_1 = \frac{7}{4}$	$M_1 = \frac{15}{8}$	$M_1 = \frac{31}{16}$

Table 5.8.15: Midpoint Rule.

- For T_1 and S_2 ,

f	g	h
$T_1 = \frac{3}{2}$	$T_1 = \frac{3}{2}$	$T_1 = \frac{3}{2}$
$S_2 = \frac{5}{3} \approx 1.6667$	$S_2 = \frac{7}{4}$	$S_2 = \frac{43}{24} \approx 1.79167$

Table 5.8.16: Trapezoid and Simpson's Rule.

d.

$$\int_0^1 f(x)dx = \frac{5}{3} \quad \int_0^1 g(x)dx = \frac{7}{4} \quad \int_0^1 h(x)dx = \frac{9}{5}$$

- e. Left endpoint rule results are overestimates; right endpoint rules are underestimates; midpoint rules are overestimates; trapezoid rules are underestimates. Simpson's rule is exact for both f and g , while a slight overestimate of $\int_0^1 h(x)dx$.

Solution.

- a. For the left and right endpoint rules, we see that

f	g	h
$L_1 = 2$	$L_1 = 2$	$L_1 = 2$
$R_1 = 1$	$R_1 = 1$	$R_1 = 1$

Table 5.8.17: Left and Trapezoid rules.

Thus, we observe that despite the fact the functions are all different, the values of L_1 and R_1 are the same for all three.

- b. For the midpoint rule, we find that

f	g	h
$M_1 = \frac{7}{4}$	$M_1 = \frac{15}{8}$	$M_1 = \frac{31}{16}$

Table 5.8.18: Midpoint Rule.

- c. For the trapezoid rule and Simpson's rule,

f	g	h
$T_1 = \frac{3}{2}$	$T_1 = \frac{3}{2}$	$T_1 = \frac{3}{2}$
$S_2 = \frac{5}{3} \approx 1.6667$	$S_2 = \frac{7}{4}$	$S_2 = \frac{43}{24} \approx 1.79167$

Table 5.8.19: Trapezoid and Simpson's Rule.

- d. The exact values of the three definite integrals are

$$\int_0^1 f(x)dx = \frac{5}{3} \quad \int_0^1 g(x)dx = \frac{7}{4} \quad \int_0^1 h(x)dx = \frac{9}{5}$$

$$\approx 1.6667 \qquad \qquad \qquad = 1.75 \qquad \qquad \qquad = 1.8$$

- e. We observe that each of the left endpoint rule results are overestimates, each of the right endpoint rules are underestimates, each of the midpoint rules are overestimates, and each of the trapezoid rules are underestimates. These results hold because each of the three functions are both decreasing and concave down. For Simpson's rule, we see that the result is exact for both f and g , while Simpson's rule is a slight overestimate of $\int_0^1 h(x)dx$.

The results seen in Example 5.8.12 generalize nicely. For instance, if f is decreasing on $[a, b]$, L_n will over-estimate the exact value of $\int_a^b f(x)dx$, and if f is concave down on $[a, b]$, M_n will over-estimate the exact value of the integral. An excellent exercise is to write a collection of scenarios of possible function behavior, and then categorize whether each of L_n , R_n , T_n , and M_n is an over- or under-estimate.

Finally, we make two important notes about Simpson's Rule. When T. Simpson first developed this rule, his idea was to replace the function f on a given interval with a quadratic function that shared three values with the function f . In so doing, he guaranteed that this new approximation rule would be exact for the definite integral of any quadratic polynomial. In one of the pleasant surprises of numerical analysis, it turns out that even though it was designed to be exact for quadratic polynomials, Simpson's Rule is exact for any cubic polynomial: that is, if we are interested in an integral such as $\int_2^5 (5x^3 - 2x^2 + 7x - 4)dx$, S_{2n} will always be exact, regardless of the value of n . This is just one more piece of evidence that shows how effective Simpson's Rule is as an approximation tool for estimating definite integrals.⁴

5.8.5 Summary

- For a definite integral such as $\int_0^1 e^{-x^2} dx$ when we cannot use the First Fundamental Theorem of Calculus because the integrand lacks an elementary algebraic antiderivative, we can estimate the integral's value by using a sequence of Riemann sum approximations. Typically, we start by computing L_n , R_n , and M_n for one or more chosen values of n .
- The Trapezoid Rule, which estimates $\int_a^b f(x)dx$ by using trapezoids, rather than rectangles, can also be viewed as the average of Left and Right Riemann sums. That is, $T_n = \frac{1}{2}(L_n + R_n)$.
- The Midpoint Rule is typically twice as accurate as the Trapezoid Rule, and the signs of the respective errors of these rules are opposites. Hence, by taking the weighted average $S_{2n} = \frac{2M_n + T_n}{3}$, we can build a much more accurate approximation to $\int_a^b f(x)dx$ by using approximations we have already computed. The rule for S_{2n} is known as Simpson's Rule, which can also be developed by approximating a given continuous function with pieces of quadratic polynomials.

⁴One reason that Simpson's Rule is so effective is that S_{2n} benefits from using $2n + 1$ points of data. Because it combines M_n , which uses n midpoints, and T_n , which uses the $n + 1$ endpoints of the chosen subintervals, S_{2n} takes advantage of the maximum amount of information we have when we know function values at the endpoints and midpoints of n subintervals.

5.8.6 Exercises

1. (Various methods for e^x numerically) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Comparison of methods for increasing concave down function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Comparing accuracy for two similar functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Identifying and comparing methods) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. Consider the definite integral $\int_0^1 x \tan(x) dx$.
 - a. Explain why this integral cannot be evaluated exactly by using either u -substitution or by integrating by parts.
 - b. Using appropriate subintervals, compute L_4 , R_4 , M_4 , T_4 , and S_8 .
 - c. Which of the approximations in (b) is an over-estimate to the true value of $\int_0^1 x \tan(x) dx$? Which is an under-estimate? How do you know?

Answer.

- a. u -substitution fails since there's not a composite function present; try showing that each of the choices of $u = x$ and $dv = \tan(x) dx$, or $u = \tan(x)$ and $dv = x dx$, fail to produce an integral that can be evaluated by parts.
- b.
 - $L_4 = 0.25892$
 - $R_4 = 0.64827$
 - $M_4 = 0.41550$
 - $T_4 = \frac{L_4+R_4}{2} = 0.45360$
 - $S_8 = \frac{2M_4+T_4}{3} = 0.42820$
- c. L_4 and M_4 are underestimates; R_4 and M_4 are overestimates.

Solution.

- a. We can't use u -substitution since there's not a function-derivative pair present (nor even a composite function). To show that integration by parts doesn't work is more complicated, but comes down to the fact that if we use the natural choice of $u = x$ and $dv = \tan(x) dx$, to do so requires us to evaluate the integral $\int \ln(\sec(x)) dx$, which appears not to have an elementary antiderivative. If we try instead $u = \tan(x)$ and $dv = x dx$, then we have to evaluate the integral $\int x^2 \cdot \sec^2(x) dx$ which is more complicated than the integral we started with and would seem to send us back where we started if we attempt integration by parts.
- b. We use computational technology to generate the following estimates, which are reported to 5 decimal places of accuracy:
 - $L_4 = 0.25892$
 - $R_4 = 0.64827$
 - $M_4 = 0.41550$
 - $T_4 = \frac{L_4+R_4}{2} = 0.45360$

- $S_8 = \frac{2M_4+T_4}{3} = 0.42820$
- c. From the values of these estimates or from the graph of $f(x) = x \tan(x)$, we can observe that f is always increasing and concave up on the interval. Thus, L_4 and M_4 are underestimates while R_4 and M_4 are overestimates.
6. For an unknown function $f(x)$, the following information is known.
- f is continuous on $[3, 6]$;
 - f is either always increasing or always decreasing on $[3, 6]$;
 - f has the same concavity throughout the interval $[3, 6]$;
 - As approximations to $\int_3^6 f(x) dx$, $L_4 = 7.23$, $R_4 = 6.75$, and $M_4 = 7.05$.
- Is f increasing or decreasing on $[3, 6]$? What data tells you?
 - Is f concave up or concave down on $[3, 6]$? Why?
 - Determine the best possible estimate you can for $\int_3^6 f(x) dx$, based on the given information.

Answer.

- Decreasing.
- Concave down.
- $\int_3^6 f(x) \approx 7.03$.

Solution.

- Because $L_4 = 7.23 > 6.75 = R_4$ and we know the function is always increasing or always decreasing, for the left endpoint rule to give a value greater than the right endpoint rule, f must be always decreasing.
- From the given data, we know that

$$T_4 = \frac{L_4 + R_4}{2} = \frac{7.23 + 6.75}{2} = 6.99.$$

Since $M_4 = 7.05$, it follows that

$$R_4 < T_4 < M_4 < L_4.$$

In particular, we know that the function is not only always decreasing on the interval, but it also concave down. Because the concavity is constant, it follows that if the trapezoid rule under-estimates the true value, then the midpoint rule must over-estimate the true value (or vice-versa). Since $T_4 < M_4$, this means that

$$T_4 < \int_3^6 f(x) dx < M_4$$

and that f must be concave down on the interval.

- We use Simpson's Rule to determine the best possible estimate for the definite integral. That is,

$$\int_3^6 f(x) dx \approx S_8$$

$$\begin{aligned}
 &= \frac{2M_4 + T_4}{3} \\
 &= \frac{2 \cdot 7.05 + 6.99}{3} \\
 &= 7.03.
 \end{aligned}$$

7. The rate at which water flows through Table Rock Dam on the White River in Branson, MO, is measured in thousands of cubic feet per second (TCFS). As engineers open the flood-gates, flow rates are recorded according to the following chart.

seconds, t	0	10	20	30	40	50	60
flow in TCFS, $r(t)$	2000	2100	2400	3000	3900	5100	6500

Table 5.8.20: Water flow data.

- What definite integral measures the total volume of water to flow through the dam in the 60 second time period provided by the table above?
- Use the given data to calculate M_n for the largest possible value of n to approximate the integral you stated in (a). Do you think M_n over- or under-estimates the exact value of the integral? Why?
- Approximate the integral stated in (a) by calculating S_n for the largest possible value of n , based on the given data.
- Compute $\frac{1}{60}S_n$ and $\frac{2000+2100+2400+3000+3900+5100+6500}{7}$. What quantity do both of these values estimate? Which is a more accurate approximation?

Answer.

- $\int_0^{60} r(t) dt$.
- $\int_0^{60} r(t) dt > M_3 = 204000$.
- $\int_0^{60} r(t) dt \approx S_6 = \frac{619000}{3} \approx 206333.33$.
- $\frac{1}{60}S_6 \approx 3438.89$; $\frac{2000+2100+2400+3000+3900+5100+6500}{7} = \frac{25000}{7} \approx 3571.43$. each estimates the average rate at which water flows through the dam on $[0, 60]$, and the first is more accurate.

Solution.

- Because $r(t)$ is measured in thousands of cubic feet per second and Δt is measured in seconds, the definite integral

$$\int_0^{60} r(t) dt$$

measures the total amount of water (in thousands of cubic feet) that flow through the dam in 60 seconds.

- With 7 function values and thus 6 subintervals as the maximum number of intervals we can use, we can compute M_3 since we need to use every other data point to evaluate

the function. Doing so,

$$\begin{aligned} \int_0^{60} r(t) dt &\approx M_3 \\ &= r(10) \cdot 20 + r(30) \cdot 20 + r(50) \cdot 20 \\ &= (2100 + 3000 + 5100) \cdot 20 \\ &= 204000. \end{aligned}$$

Thus, approximately 204000000 cubic feet of water pass through the dam on the interval $0 \leq t \leq 60$.

We observe that r appears to be increasing at an increasing rate on $0 \leq t \leq 60$. In particular, the data suggests that r is concave up on the interval. As such, the trapezoid rule will over-estimate the true value of the definite integral and the midpoint rule will under-estimate the value. Thus,

$$\int_0^{60} r(t) dt > M_3 = 204000.$$

- c. In order to estimate the integral using S_6 , we need to know the value of T_3 along with M_3 . To find T_3 , we compute L_3 and R_3 . First

$$L_3 = r(0) \cdot 20 + r(20) \cdot 20 + r(40) \cdot 20 = (2000 + 2400 + 3900) \cdot 20 = 166000.$$

Similarly,

$$R_3 = r(20) \cdot 20 + r(40) \cdot 20 + r(60) \cdot 20 = (2400 + 3900 + 6500) \cdot 20 = 256000.$$

Thus,

$$T_3 = \frac{L_3 + R_3}{2} = \frac{166000 + 256000}{2} = 211000.$$

Finally, we now have the information needed to compute S_6 . We thus find that

$$\int_0^{60} r(t) dt \approx S_6 = \frac{2M_3 + T_3}{3} = \frac{2 \cdot 204000 + 211000}{3} = \frac{619000}{3} \approx 206333.33.$$

- d. First,

$$\frac{1}{60} S_6 = \frac{206333.33}{60} = \frac{30950}{9} \approx 3438.89.$$

This is one estimate for the average value of r on $[0, 60]$ since

$$r_{\text{AVG}[0,60]} = \frac{1}{60 - 0} \int_0^{60} r(t) dt,$$

and $\int_0^{60} r(t) dt \approx S_6$. Thus, on average, water is flowing through the dam at a rate of 3438.89 thousand cubic feet per second.

Next,

$$\frac{2000 + 2100 + 2400 + 3000 + 3900 + 5100 + 6500}{7} = \frac{25000}{7} \approx 3571.43$$

is the average of the 7 known values of $r(t)$ on the interval $[0, 60]$, and thus is one estimate of the average value of r . This average treats all of them equally, as opposed to S_6 which counts midpoints more than endpoints and is developed in such a way to remove some of the error inherent in using simple estimates like L_6 and R_6 . This simple average of 7 rate values is analogous to using a left or right Riemann sum, and thus we expect that the estimate of the average value of r that comes from S_6 is more accurate for computing the average rate at which water flows through the dam.

Chapter 6

Using Definite Integrals

6.1 Using Definite Integrals to Find Area and Length

Motivating Questions

- 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?
 - How can a definite integral be used to measure the length of a curve?
-

Early on in our work with the definite integral, we learned that for an object moving along an axis, the area under a non-negative velocity function v between a and b tells us the distance the object traveled on that time interval, and that area is given precisely by the definite integral $\int_a^b v(t) dt$. In general, for any nonnegative function f on an interval $[a, b]$, $\int_a^b f(x) dx$ measures the area bounded by the curve and the x -axis between $x = a$ and $x = b$.

Next, we will explore how definite integrals can be used to represent other physically important properties. In Example 6.1.1, we investigate how a single definite integral may be used to represent the area between two curves.

Example 6.1.1. Consider the functions given by $f(x) = 5 - (x - 1)^2$ and $g(x) = 4 - x$.

- a. Use algebra to find the points where the graphs of f and g intersect.
- b. Sketch an accurate graph of f and g on the axes provided, labeling the curves by name and the intersection points with ordered pairs.
- c. Find and evaluate exactly an integral expression that represents the area between $y = f(x)$ and the x -axis on the interval between the intersection points of f and g .

- d. Find and evaluate exactly an integral expression that represents the area between $y = g(x)$ and the x -axis on the interval between the intersection points of f and g .
- e. What is the exact area between f and g between their intersection points? Why?

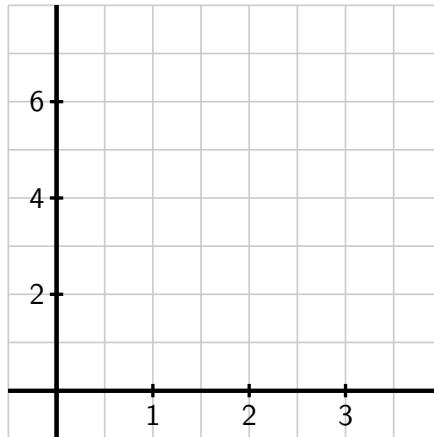


Figure 6.1.2: Axes for plotting f and g in Example 6.1.1

6.1.1 The Area Between Two Curves

In Example 6.1.1, we saw a natural way to think about the area between two curves: it is the area beneath the upper curve minus the area below the lower curve.

Example 6.1.3. Find the area bounded between the graphs of $f(x) = (x - 1)^2 + 1$ and $g(x) = x + 2$.

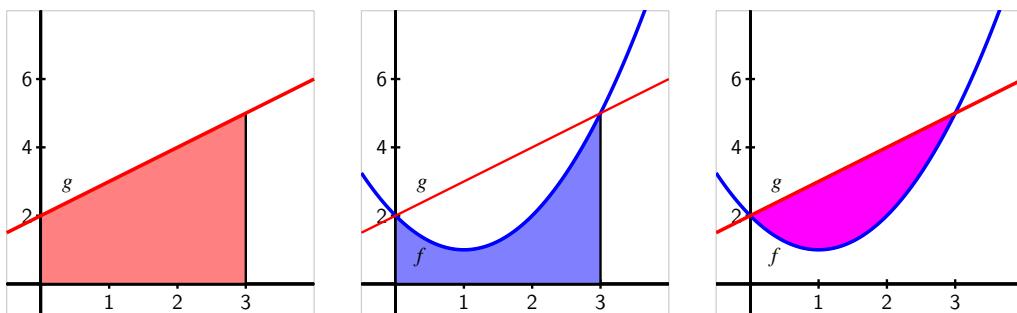


Figure 6.1.4: The areas bounded by the functions $f(x) = (x - 1)^2 + 1$ and $g(x) = x + 2$ on the interval $[0, 3]$.

Solution. In Figure 6.1.4, we see that the graphs intersect at $(0, 2)$ and $(3, 5)$. We can find

these intersection points algebraically by solving the system of equations given by $y = x + 2$ and $y = (x - 1)^2 + 1$: substituting $x + 2$ for y in the second equation yields $x + 2 = (x - 1)^2 + 1$, so $x + 2 = x^2 - 2x + 1 + 1$, and thus

$$x^2 - 3x = x(x - 3) = 0,$$

from which it follows that $x = 0$ or $x = 3$. Using $y = x + 2$, we find the corresponding y -values of the intersection points.

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

$$A = \int_0^3 (x + 2) dx - \int_0^3 [(x - 1)^2 + 1] dx = \frac{21}{2} - 6 = \frac{9}{2}. \quad (6.1.1)$$

We can also think of the area this way: if we slice up the region between two curves 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), we see (as shown in Figure 6.1.5) that the height of a typical rectangle is given by the difference between the two functions, $g(x) - f(x)$, and its width is Δx . Thus the area of the rectangle is

$$A_{\text{rect}} = (g(x) - f(x))\Delta x.$$

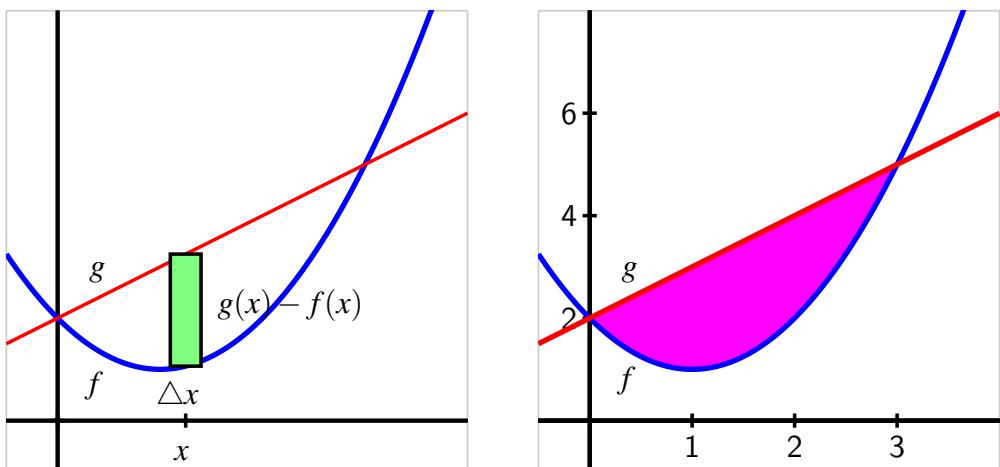


Figure 6.1.5: The area bounded by the functions $f(x) = (x - 1)^2 + 1$ and $g(x) = x + 2$ on the interval $[0, 3]$.

The area between the two curves on $[0, 3]$ is thus approximated by the Riemann sum

$$A \approx \sum_{i=1}^n (g(x_i) - f(x_i))\Delta x,$$

and as we let $n \rightarrow \infty$, it follows that the area is given by the single definite integral

$$A = \int_0^3 (g(x) - f(x)) dx. \quad (6.1.2)$$

In many applications of the definite integral, we will find it helpful to think of a “representative slice” and use the definite integral to add these slices. Here, the integral sums the areas of thin rectangles.

Finally, it doesn’t matter whether we think of the area between two curves as the difference between the area bounded by the individual curves (as in (6.1.1)) or as the limit of a Riemann sum of the areas of thin rectangles between the curves (as in (6.1.2)). These two results are the same, since the difference of two integrals is the integral of the difference:

$$\int_0^3 g(x) dx - \int_0^3 f(x) dx = \int_0^3 (g(x) - f(x)) dx.$$

Our work so far in this section illustrates the following general principle.

If two curves $y = g(x)$ and $y = f(x)$ intersect at $(a, g(a))$ and $(b, g(b))$, and for all x such that $a \leq x \leq b$, $g(x) \geq f(x)$, then the area between the curves is $A = \int_a^b (g(x) - f(x)) dx$.

Example 6.1.6. 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.

- The finite region bounded by $y = \sqrt{x}$ and $y = \frac{1}{4}x$.
- The finite region bounded by $y = 12 - 2x^2$ and $y = x^2 - 8$.
- 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.
- The finite regions between the curves $y = x^3 - x$ and $y = x^2$.

Hint.

- Find where the two curves intersect.

- b. Plot both functions and think about how they compare to $y = x^2$.
- c. Remember that $\sin(\pi/4) = \cos(\pi/4)$.
- d. Find three points where the curves intersect and thus two finite regions between two pairs of those points.

Answer.

a. $A = \int_0^{16} (\sqrt{x} - \frac{1}{4}x) dx = \frac{32}{3}$.

b. $A = \int_{-\sqrt{20}/3}^{\sqrt{20}/3} ((12 - x^2) - (x^2 - 8)) dx = \frac{160\sqrt{\frac{5}{3}}}{3} \approx 68.853$.

c. $A = \int_0^{\frac{\pi}{4}} (\cos(x) - \sin(x)) dx = \sqrt{2} - 1$.

d. The left-hand region has area

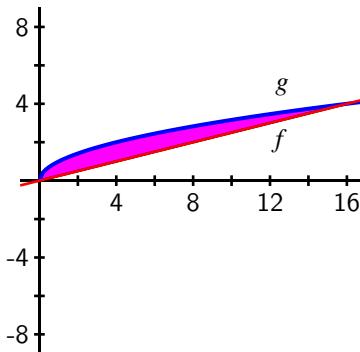
$$A_1 = \int_{\frac{1-\sqrt{5}}{2}}^0 (x^2 - (x^3 - x)) dx = \frac{13 + 5\sqrt{5}}{24} \approx 1.007514.$$

The right-hand region has area

$$A_2 = \int_0^{\frac{1+\sqrt{5}}{2}} ((x^3 - x) - x^2) dx = \frac{13 - 5\sqrt{5}}{24} \approx 0.075819.$$

Solution.

- a. By solving $\sqrt{x} = \frac{1}{4}x$, we see that the two curves intersect when $x = 0$ and $x = 16$, so at the points $(0, 0)$ and $(16, 4)$. We can plot the region as shown in the figure below.

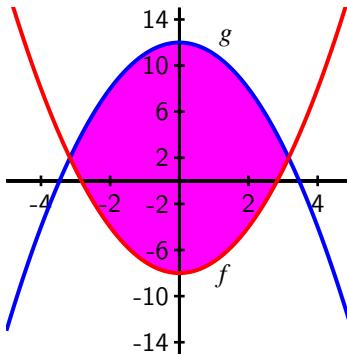


Slicing the region with thin vertical rectangles, we see that a typical rectangle's height is given by $\sqrt{x} - \frac{1}{4}x$, and thus the area of such a slice is $A_{\text{rect}} = (\sqrt{x} - \frac{1}{4}x)\Delta x$. It follows that the area of the region bounded by the two curves is

$$A = \int_0^{16} (\sqrt{x} - \frac{1}{4}x) dx.$$

Evaluating the integral, we find $A = \frac{32}{3}$.

- b. Solving $12 - 2x^2 = x^2 - 8$, we find that the two curves intersect where $3x^2 = 20$, so at $x = \pm\sqrt{20/3}$. We can plot the region as shown in the figure below.

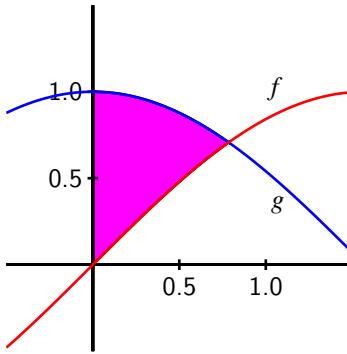


Slicing the region with thin vertical rectangles, a typical rectangle's height is given by $(12 - x^2) - (x^2 - 8)$, and thus the area of such a slice is $A_{\text{rect}} = ((12 - x^2) - (x^2 - 8))\Delta x$. Hence, the area of the region bounded by the two curves is

$$A = \int_{-\sqrt{20/3}}^{\sqrt{20/3}} ((12 - x^2) - (x^2 - 8)) dx.$$

Evaluating the integral, we find $A = \frac{160\sqrt{5}}{3} \approx 68.853$.

- c. The first positive value at which the sine and cosine functions intersect is $x = \frac{\pi}{4}$. Plotting the functions in a surrounding region, we see the following figure.



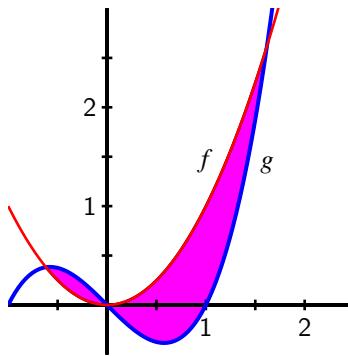
Slicing the region with vertical rectangles, we find the area of the shaded region is given by

$$A = \int_0^{\pi/4} (\cos(x) - \sin(x)) dx = \sqrt{2} - 1.$$

- d. The curves $y = x^3 - x$ and $y = x^2$ intersect at three points, forming two regions. Setting $x^3 - x = x^2$ and solving, we find that

$$x^3 - x^2 - x = x(x^2 - x - 1) = 0,$$

which implies that $x = 0$ or $x = \frac{1}{2}(1 \pm \sqrt{5})$. Plotting the curves, we see that for the left-hand region, the relative interval is $[\frac{1-\sqrt{5}}{2}, 0]$ and $g(x) = x^3 - x > x^2 = f(x)$.



Thus, the left-hand region has area

$$A_1 = \int_{\frac{1-\sqrt{5}}{2}}^0 (x^2 - (x^3 - x)) dx = \frac{13 + 5\sqrt{5}}{24} \approx 1.007514.$$

For the right-hand region, the interval is $[0, \frac{1+\sqrt{5}}{2}]$ and $f(x) = x^2 > x^3 - x = g(x)$, so that its area is

$$A_2 = \int_0^{\frac{1+\sqrt{5}}{2}} ((x^3 - x) - x^2) dx = \frac{13 - 5\sqrt{5}}{24} \approx 0.075819.$$

6.1.2 Finding Area with Horizontal Slices

At times, the shape of a region may dictate that we use horizontal rectangular slices, instead of vertical ones.

Example 6.1.7. Find the area of the region bounded by the parabola $x = y^2 - 1$ and the line $y = x - 1$, shown at left in Figure 6.1.8.

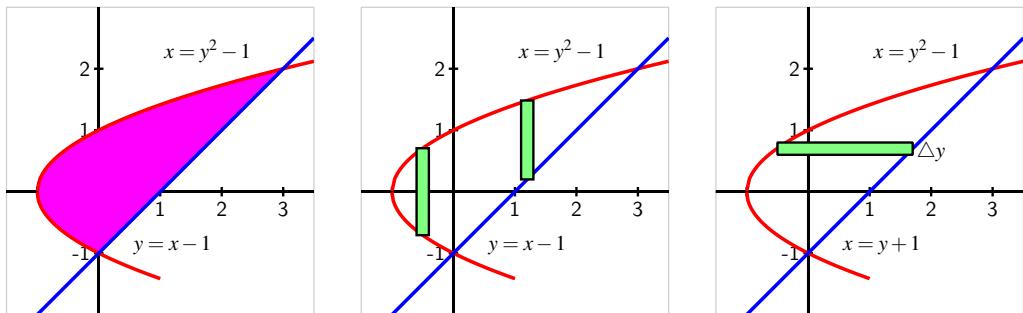


Figure 6.1.8: The area bounded by the functions $x = y^2 - 1$ and $y = x - 1$ (at left), with the region sliced vertically (center) and horizontally (at right).

Solution. By solving the second equation for x and writing $x = y + 1$, we find that $y + 1 = y^2 - 1$. Hence the curves intersect where $y^2 - y - 2 = 0$. Thus, we find $y = -1$ or $y = 2$, so the intersection points of the two curves are $(0, -1)$ and $(3, 2)$.

If we attempt to use vertical rectangles to slice up the area (as in the center graph of Figure 6.1.8), we see that from $x = -1$ to $x = 0$ the curves that bound the top and bottom of the rectangle are one and the same. This suggests, as shown in the rightmost graph in the figure, that we try using horizontal rectangles.

Note that the width of a horizontal rectangle depends on y . Between $y = -1$ and $y = 2$, the right end of a representative rectangle is determined by the line $x = y + 1$, and the left end is determined by the parabola, $x = y^2 - 1$. The thickness of the rectangle is Δy .

Therefore, the area of the rectangle is

$$A_{\text{rect}} = [(y + 1) - (y^2 - 1)]\Delta y,$$

and the area between the two curves on the y -interval $[-1, 2]$ is approximated by the Riemann sum

$$A \approx \sum_{i=1}^n [(y_i + 1) - (y_i^2 - 1)]\Delta y.$$

Taking the limit of the Riemann sum, it follows that the area of the region is

$$A = \int_{y=-1}^{y=2} [(y + 1) - (y^2 - 1)] dy. \quad (6.1.3)$$

We emphasize that we are integrating with respect to y ; this is because we chose to use horizontal rectangles whose widths depend on y and whose thickness is denoted Δy . It is a straightforward exercise to evaluate the integral in Equation (6.1.3) and find that $A = \frac{9}{2}$.

Just as with the use of vertical rectangles of thickness Δx , we have a general principle for finding the area between two curves, which we state as follows.

If two curves $x = g(y)$ and $x = f(y)$ intersect at $(g(c), c)$ and $(g(d), d)$, and for all y such that $c \leq y \leq d$, $g(y) \geq f(y)$, then the area between the curves is

$$A = \int_{y=c}^{y=d} (g(y) - f(y)) dy.$$

Example 6.1.9. 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.

- 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 $y = x$.

Hint.

- a. Find where the curves intersect and plot them.
- b. Observe that horizontal slices are appropriate, so we will integrate with respect to y .
- c. Note that the region is approximately triangular with the x -axis as the base, $y = x^2$ as the left "side", and $y = 2 - x$ as the right side.
- d.
- e. Plot the region and observe that horizontal slices are appropriate because of the shape of the region.

Answer.

- a. $A = \int_{y=-\sqrt{2}}^{y=\sqrt{2}} (6 - 2y^2 - y^2) dy = 8\sqrt{2} \approx 11.314.$
- b. $A = \int_{y=-1}^{y=1} (2 - 2y^2 - (1 - y^2)) dy = \frac{4}{3}.$
- c. $A = \int_{y=0}^{y=4-2\sqrt{3}} \left(\frac{2-y}{2} - \sqrt{y}\right) dy = \frac{11}{3} - 2\sqrt{3} \approx 0.2026$
- d. $A = \int_0^3 (y - (y^2 - 2y)) dy = \frac{9}{2}.$

Solution.

- a. The two curves intersect when $y^2 = 6 - 2y^2$, thus when $3y^2 = 6$, so at $y = \pm\sqrt{2}$. Plotting the region, we see that for $y = -\sqrt{2}$ to $y = \sqrt{2}$, $g(y) = 6 - 2y^2 > y^2 = f(y)$, and thus horizontal slices are appropriate for setting up the integral that represents the region's area. The area is thus given by

$$A = \int_{y=-\sqrt{2}}^{y=\sqrt{2}} (6 - 2y^2 - y^2) dy = 8\sqrt{2} \approx 11.314.$$

- b. The two curves intersect when $1 - y^2 = 2 - 2y^2$, and therefore $3y^2 = 3$ so $y = \pm 1$. Graphing the region, on $-1 < y < 1$, we see that $g(y) = 2 - 2y^2 \geq 1 - y^2 = f(y)$. Using horizontal slices, it follows that the region's area is given by

$$A = \int_{y=-1}^{y=1} (2 - 2y^2 - (1 - y^2)) dy = \int_{y=-1}^{y=1} (1 - y^2) dy = \frac{4}{3}.$$

- c. Standard algebra shows that the two curves meet at $(-1 + \sqrt{3}, 4 - 2\sqrt{3})$. A careful plot of the region shows that the two curves, together with the x -axis, form a roughly triangular region with the x -axis as the base, $y = x^2$ as the left "side", and $y = 2 - 2x$ as the right side. Using horizontal slices, and writing each respective equation with x as a function of y (so $x = \sqrt{y}$ and $x = \frac{2-y}{2}$), we find that the area of this region is

$$A = \int_{y=0}^{y=4-2\sqrt{3}} \left(\frac{2-y}{2} - \sqrt{y} \right) dy = \frac{11}{3} - 2\sqrt{3} \approx 0.2026$$

- d. The two curves intersect at $(0, 0)$ and $(3, 3)$. For $0 \leq y \leq 3$, we observe that $g(y) = y \geq y^2 - 2y = f(y)$, so using horizontal slices and the corresponding definite integral, we find that the region's area is

$$A = \int_0^3 (y - (y^2 - 2y)) dy = \int_0^3 (3y - y^2) dy = \frac{9}{2}.$$

6.1.3 Finding the length of a curve

We can also use the definite integral to find the length of a portion of a curve. We use the same fundamental principle: we slice the curve up into small pieces whose lengths we can easily approximate. Specifically, we subdivide the curve into small approximating line segments, as shown at left in Figure 6.1.10.

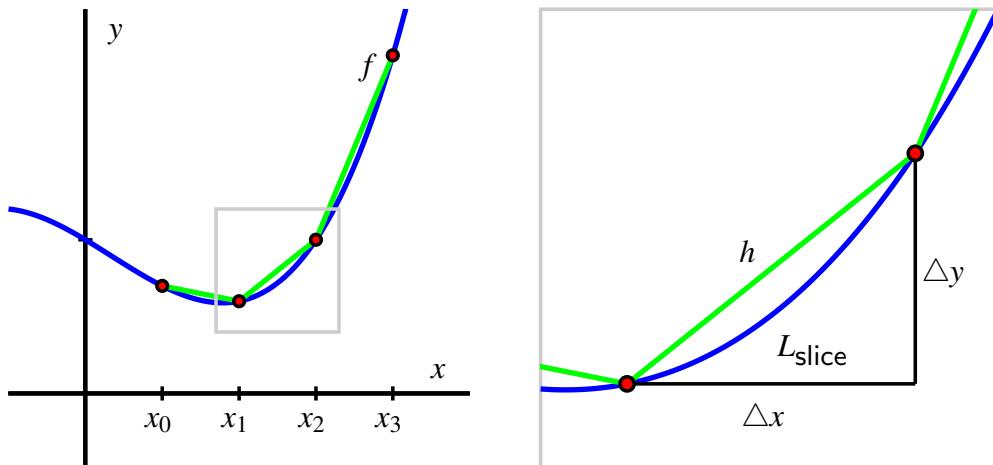


Figure 6.1.10: At left, a continuous function $y = f(x)$ whose length we seek on the interval $a = x_0$ to $b = x_3$. At right, a close up view of a portion of the curve.

We estimate the length L_{slice} of each portion of the curve on a small interval of length Δx . We use the right triangle with legs parallel to the coordinate axes and hypotenuse connecting

the endpoints of the slice, as seen at right in Figure 6.1.10. The length, h , of the hypotenuse approximates the length, L_{slice} , of the curve between the two selected points. Thus,

$$L_{\text{slice}} \approx h = \sqrt{(\Delta x)^2 + (\Delta y)^2}.$$

Next we use algebra to rearrange the expression for the length of the hypotenuse into a form that we can integrate. By removing a factor of $(\Delta x)^2$, we find

$$\begin{aligned} L_{\text{slice}} &\approx \sqrt{(\Delta x)^2 + (\Delta y)^2} \\ &= \sqrt{(\Delta x)^2 \left(1 + \frac{(\Delta y)^2}{(\Delta x)^2}\right)} \\ &= \sqrt{1 + \frac{(\Delta y)^2}{(\Delta x)^2}} \cdot \Delta x. \end{aligned}$$

Then, as $n \rightarrow \infty$ and $\Delta x \rightarrow 0$, we have that $\frac{\Delta y}{\Delta x} \rightarrow \frac{dy}{dx} = f'(x)$. Thus, we can say that

$$L_{\text{slice}} \approx \sqrt{1 + f'(x)^2} \Delta x.$$

Taking a Riemann sum of all of these slices and letting $n \rightarrow \infty$, we arrive at the following fact.

Given a differentiable function f on an interval $[a, b]$, the total arc length, L , along the curve $y = f(x)$ from $x = a$ to $x = b$ is given by

$$L = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

Example 6.1.11. Each of the following questions somehow involves the arc length along a curve.

- Use the definition and appropriate computational technology to determine the arc length along $y = x^2$ from $x = -1$ to $x = 1$.
- Find the arc length of $y = \sqrt{4 - x^2}$ on the interval $-2 \leq x \leq 2$. Find this value in two different ways: (a) by using a definite integral, and (b) by using a familiar property of the curve.
- Determine the arc length of $y = xe^{3x}$ on the interval $[0, 1]$.
- Will the integrals that arise calculating arc length typically be ones that we can evaluate exactly using the First FTC, or ones that we need to approximate? Why?
- A moving particle is traveling along the curve given by $y = f(x) = 0.1x^2 + 1$, and does so at a constant rate of 7 cm/sec, where both x and y are measured in cm (that is, the

curve $y = f(x)$ is the path along which the object actually travels; the curve is not a “position function”). Find the position of the particle when $t = 4$ sec, assuming that when $t = 0$, the particle’s location is $(0, f(0))$.

Hint.

- Recall that $L = \int_a^b \sqrt{1 + f'(x)^2} dx$.
- Remember that $x^2 + y^2 = 4$ generates a circle centered at $(0, 0)$ of radius 4.
- Apply the arc length formula. Use technology to evaluate the integral.
- Most expressions involving square roots are difficult to antiderivative.
- Here you can determine the arc length and then experiment to find the upper limit of integration, which will help you determine position.

Answer.

- $L \approx 2.95789$.
- $L = \int_{-2}^2 \sqrt{\frac{4}{4-x^2}} dx = 2\pi$.
- $L = \int_0^1 \sqrt{1 + e^{6x}(9x^2 + 6x + 1)} dx \approx 20.1773$.
- We will usually have to estimate the value of $\int_a^b \sqrt{1 + f'(x)^2} dx$ using computational technology.
- Approximately $(9.011, f(9.011)) = (9.011, 9.1198)$.

Solution.

- Using the formula for arc length, we know

$$L = \int_{-1}^1 \sqrt{1 + (2x)^2} dx.$$

Evaluating the integral, we find $L \approx 2.95789$.

- The curve is a semi-circle of radius 2, so its length must be $L = \frac{1}{2} \cdot 2\pi r = 2\pi$. This is confirmed by the arc length formula, since $f(x) = \sqrt{4 - x^2}$, so $f'(x) = \frac{1}{2}(4 - x^2)^{-1/2}(-2x)$

$$L = \int_{-2}^2 \sqrt{1 + (-x(4 - x^2)^{-1/2})^2} dx = \int_{-2}^2 \sqrt{1 + x^2(4 - x^2)^{-1}} dx.$$

It follows¹ that

$$L = \int_{-2}^2 \sqrt{\frac{4}{4 - x^2}} dx = 2\pi.$$

¹This integral is actually “improper” because the integrand is undefined at the endpoints, $x = \pm 2$. We will learn how to evaluate such integrals in Section 6.5.

- c. For $y = xe^{3x}$, we observe that $y' = 3xe^{3x} + e^{3x} = e^{3x}(3x + 1)$, and that

$$(y')^2 = e^{6x}(9x^2 + 6x + 1).$$

Thus, the arc length of the curve on $[0, 1]$ is

$$L = \int_0^1 \sqrt{1 + e^{6x}(9x^2 + 6x + 1)} dx \approx 20.1773.$$

- d. Because $\sqrt{1 + f'(x)^2}$ will rarely simplify nicely and rarely have an elementary antiderivative, we will usually have to estimate the value of $\int_a^b \sqrt{1 + f'(x)^2} dx$ using computational technology.
- e. After 4 seconds traveling at 7 cm/sec, the particle has moved 28 cm. So, we have to find the value of b for which the arc length L along $[0, b]$ is 28. Thus, we know that

$$28 = L = \int_0^b \sqrt{1 + (0.2x)^2} dx.$$

Rewriting the integral, we see

$$28 = \int_0^b \sqrt{1 + 0.4x^2} dx.$$

Experimenting with different values of b in a computational engine like Wolfram | Alpha, we find that for $b \approx 9.011$, $\int_0^{9.011} \sqrt{1 + 0.4x^2} dx \approx 27.9992$, and thus the position of the particle when $t = 4$ is approximately $(9.011, f(9.011)) = (9.011, 9.1198)$.

6.1.4 Summary

- 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 $A_{\text{rect}} = (g(x) - f(x))\Delta x$, 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 Δx or horizontal rectangles of thickness Δy . We want the height of the rectangle given 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.
- The arc length, L , along the curve $y = f(x)$ from $x = a$ to $x = b$ is given by

$$L = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

6.1.5 Exercises

- (Area between two power functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Area between two trigonometric functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Area between two curves) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Arc length of a curve) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- Find the exact area of each described region.
 - The finite region between the curves $x = y(y - 2)$ and $x = -(y - 1)(y - 3)$.
 - The region between the sine and cosine functions on the interval $[\frac{\pi}{4}, \frac{3\pi}{4}]$.
 - The finite region between $x = y^2 - y - 2$ and $y = 2x - 1$.
 - The finite region between $y = mx$ and $y = x^2 - 1$, where m is a positive constant.

Answer.

- $A = \int_{\frac{3-\sqrt{3}}{2}}^{\frac{3+\sqrt{3}}{2}} -2y^2 + 6y - 3 \, dy = \sqrt{3}$.
- $A = \int_{\pi/4}^{3\pi/4} \sin(x) - \cos(x) \, dx = \sqrt{2}$.
- $A = \int_{-1}^{5/2} \frac{y+1}{2} - (y^2 - y - 2) \, dy = \frac{343}{48}$.
- $A = \int_{\frac{m-\sqrt{m^2+4}}{2}}^{\frac{m+\sqrt{m^2+4}}{2}} mx - (x^2 - 1) \, dx = -\frac{1}{3} \left(\frac{m+\sqrt{m^2+4}}{2} \right)^3 - \frac{1}{3} \left(\frac{m-\sqrt{m^2+4}}{2} \right)^3 + \frac{m}{2} \left(\frac{m+\sqrt{m^2+4}}{2} \right)^2 - \frac{m}{2} \left(\frac{m-\sqrt{m^2+4}}{2} \right)^2 + \left(\frac{m+\sqrt{m^2+4}}{2} - \frac{m-\sqrt{m^2+4}}{2} \right)$.

Solution.

- The point of intersection of the graphs occurs when $y(y - 2) = -(y - 1)(y - 3)$, and thus we need to find where $y^2 - 2y = -(y^2 - 4y + 3)$ or $2y^2 - 6y + 3 = 0$. By the quadratic formula, it follows that

$$y = \frac{6 \pm \sqrt{(-6)^2 - (4)(2)(3)}}{(2)(2)} = \frac{6 \pm \sqrt{36 - 24}}{4} = \frac{6 \pm \sqrt{12}}{4} = \frac{3 \pm \sqrt{3}}{2}.$$

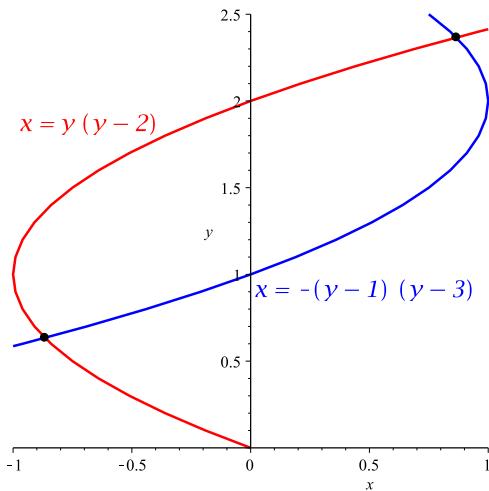
Since the curve $x = -(y - 1)(y - 3)$ lies to the right (has a greater x value) of $x = y(y - 2)$ between $y = \frac{3-\sqrt{3}}{2} \approx 0.63$ and $y = \frac{3+\sqrt{3}}{2} \approx 2.37$ as shown in the figure below, we see that the area between the two curves is given by the integral

$$\int_{\frac{3-\sqrt{3}}{2}}^{\frac{3+\sqrt{3}}{2}} -(y - 1)(y - 3) - y(y - 2) \, dy.$$

Evaluating this integral, we find that

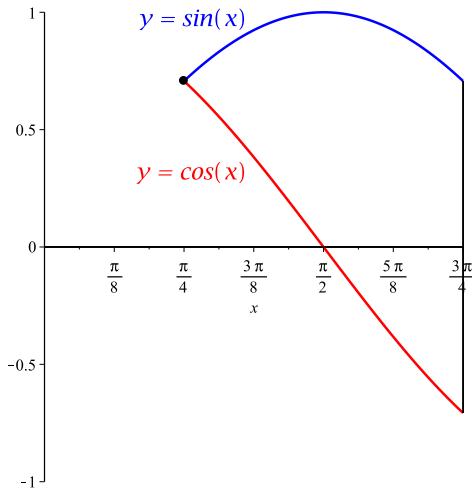
$$A = \int_{\frac{3-\sqrt{3}}{2}}^{\frac{3+\sqrt{3}}{2}} -2y^2 + 6y - 3 \, dy$$

$$\begin{aligned}
&= \left(-\frac{2}{3}y^3 + 3y^2 - 3y \right) \Big|_{\frac{3-\sqrt{3}}{2}}^{\frac{3+\sqrt{3}}{2}} \\
&= \left[-\frac{2}{3} \left(\frac{3+\sqrt{3}}{2} \right)^3 + 3 \left(\frac{3+\sqrt{3}}{2} \right)^2 - 3 \left(\frac{3+\sqrt{3}}{2} \right) \right] \\
&\quad - \left[-\frac{2}{3} \left(\frac{3-\sqrt{3}}{2} \right)^3 + 3 \left(\frac{3-\sqrt{3}}{2} \right)^2 - 3 \left(\frac{3-\sqrt{3}}{2} \right) \right] \\
&= -\frac{1}{12}(54 + 30\sqrt{3}) + \frac{1}{12}(54 - 30\sqrt{3}) + \frac{3}{4}(12 + 6\sqrt{3}) \\
&\quad - \frac{3}{4}(12 - 6\sqrt{3}) - \frac{3}{2}(3 + \sqrt{3}) + \frac{3}{2}(3 - \sqrt{3}) \\
&= -5\sqrt{3} + 9\sqrt{3} - 3\sqrt{3} \\
&= \sqrt{3}.
\end{aligned}$$



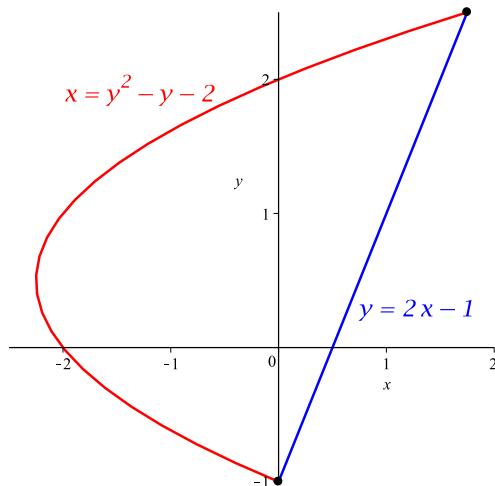
- b. As the figure below shows, the graph of $y = \sin(x)$ lies above the graph of $y = \cos(x)$ on this interval. So the area between the two curves is given by

$$\begin{aligned}
A &= \int_{\pi/4}^{3\pi/4} \sin(x) - \cos(x) \, dx \\
&= (-\cos(x) - \sin(x)) \Big|_{\pi/4}^{3\pi/4} \\
&= \left[-\cos\left(\frac{3\pi}{4}\right) - \sin\left(\frac{3\pi}{4}\right) \right] - \left[-\cos\left(\frac{\pi}{4}\right) - \sin\left(\frac{\pi}{4}\right) \right] \\
&= \left[\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2} \right] + \left[\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} \right] \\
&= \sqrt{2}.
\end{aligned}$$



- c. We can rewrite $y = 2x - 1$ as $x = \frac{y+1}{2}$. The two curves intersect when $y^2 - y - 2 = \frac{y+1}{2}$. It follows that $2y^2 - 2y - 4 = y + 1$, so $2y^2 - 3y - 5 = 0$ and thus factoring, $(2y - 5)(y + 1) = 0$. Hence the two curves intersect when $y = -1$ or $y = \frac{5}{2}$. The figure below shows that $x = \frac{y+1}{2}$ lies to the right of $x = y^2 - y - 2$ on this interval, so the area of the region between the two curves is

$$\begin{aligned}
 A &= \int_{-1}^{\frac{5}{2}} 5/2 \frac{y+1}{2} - (y^2 - y - 2) \, dy \\
 &= \int_{-1}^{\frac{5}{2}} 5/2 - y^2 + \frac{3}{2}y + \frac{5}{2} \, dy \\
 &= \left(-\frac{y^3}{3} + \frac{3}{4}y^2 + \frac{5}{2}y \right) \Big|_{-1}^{\frac{5}{2}} \\
 &= -\frac{1}{3} \left[\left(\frac{5}{2}\right)^3 + 1 \right] + \frac{3}{4} \left[\left(\frac{5}{2}\right)^2 - 1 \right] + \frac{5}{2} \left[\left(\frac{5}{2}\right) + 1 \right] \\
 &= -\frac{133}{24} + \frac{63}{16} + \frac{35}{4} \\
 &= \frac{343}{48}.
 \end{aligned}$$

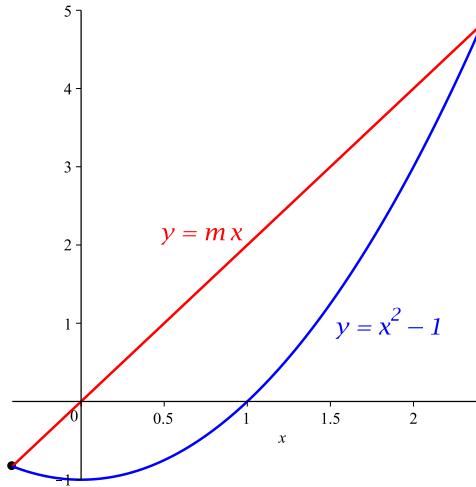


- d. A representative picture (with $m = 2$) is shown in the figure below. The two curves intersect when $x^2 - 1 = mx$, or when $x^2 - mx - 1 = 0$. By the quadratic formula, it follows that

$$x = \frac{m \pm \sqrt{m^2 - 4(1)(-1)}}{2} = \frac{m \pm \sqrt{m^2 + 4}}{2}.$$

If we integrate with vertical slices, then the graph of $y = mx$ is always above the graph of $y = x^2 - 1$ and so the area of the region between the two graphs is

$$\begin{aligned} A &= \int_{\frac{m-\sqrt{m^2+4}}{2}}^{\frac{m+\sqrt{m^2+4}}{2}} mx - (x^2 - 1) \, dx \\ &= \left(-\frac{x^3}{3} + m \frac{x^2}{2} + x \right) \Big|_{\frac{m-\sqrt{m^2+4}}{2}}^{\frac{m+\sqrt{m^2+4}}{2}} \\ &= -\frac{1}{3} \left(\frac{m + \sqrt{m^2 + 4}}{2} \right)^3 - \frac{1}{3} \left(\frac{m - \sqrt{m^2 + 4}}{2} \right)^3 + \frac{m}{2} \left(\frac{m + \sqrt{m^2 + 4}}{2} \right)^2 \\ &\quad - \frac{m}{2} \left(\frac{m - \sqrt{m^2 + 4}}{2} \right)^2 + \left(\frac{m + \sqrt{m^2 + 4}}{2} - \frac{m - \sqrt{m^2 + 4}}{2} \right). \end{aligned}$$



6. Let $f(x) = 1 - x^2$ and $g(x) = ax^2 - a$, where a is an unknown positive real number. For what value(s) of a is the area between the curves f and g equal to 2?

Answer. $a = \frac{1}{2}$.

Solution. Both of these quadratic functions pass through the points $(-1, 0)$ and $(1, 0)$, and thus these are the two intersection points for all positive values of a . In addition, since $1 - x^2 > ax^2 - a$ on $-1 < x < 1$, it follows that the area between the curves is given by

$$A = \int_{-1}^1 (1 - x^2 - (ax^2 - a)) \, dx = \int_{-1}^1 ((1 + a) - (1 + a)x^2) \, dx.$$

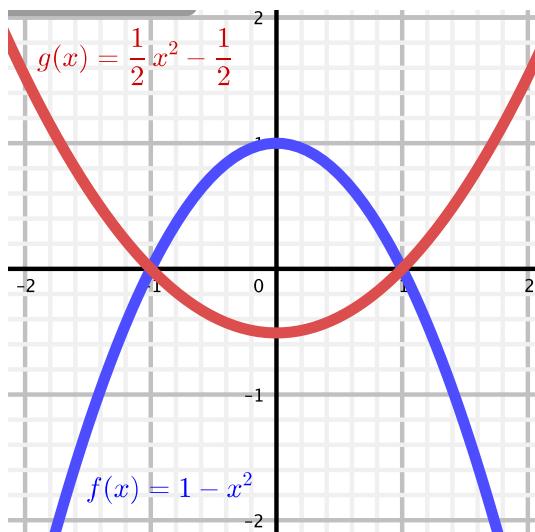
Evaluating the integral, we find that

$$\begin{aligned} A &= \int_{-1}^1 (1 + a - (1 + a)x^2) \, dx \\ &= (1 + a)x - \frac{1 + a}{3}x^3 \Big|_{-1}^1 \\ &= (1 + a)(1) - \frac{1 + a}{3}(1)^3 - \left((1 + a)(-1) - \frac{1 + a}{3}(-1)^3 \right) \\ &= 2(1 + a) - \frac{2(1 + a)}{3} \\ &= \frac{4(1 + a)}{3}. \end{aligned}$$

We want to know for which value(s) of a it follows that $A = 2$. Thus, we solve

$$2 = \frac{4(1 + a)}{3},$$

from which it follows that $a = \frac{1}{2}$. Hence, in the figure below, the area between the two curves is exactly $A = 2$.



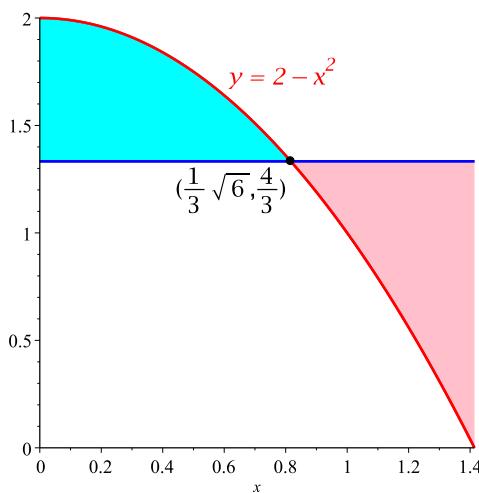
7. Let $f(x) = 2 - x^2$. Recall that the average value of any continuous function f on an interval $[a, b]$ is given by $\frac{1}{b-a} \int_a^b f(x) dx$.

- Find the average value of $f(x) = 2 - x^2$ on the interval $[0, \sqrt{2}]$. Call this value r .
- Sketch a graph of $y = f(x)$ and $y = r$. Find their intersection point(s).
- Show that on the interval $[0, \sqrt{2}]$, the amount of area that lies below $y = f(x)$ and above $y = r$ is equal to the amount of area that lies below $y = r$ and above $y = f(x)$.
- Will the result of (c) be true for any continuous function and its average value on any interval? Why?

Answer.

a. $r = \frac{4}{3}$.

b.



c. $A_1 = A_2 = \frac{4\sqrt{6}}{27}$.

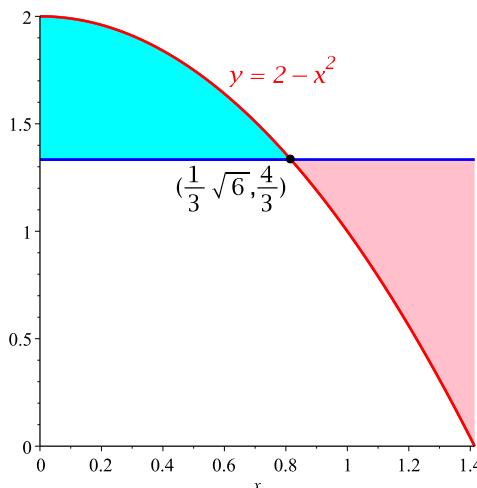
d. Yes.

Solution.

a. The average value of f on the interval $[0, \sqrt{2}]$ is

$$\begin{aligned}
 r &= \frac{1}{\sqrt{2}} \int_0^{\sqrt{2}} 2 - x^2 \, dx \\
 &= \frac{1}{\sqrt{2}} \left(2x - \frac{x^3}{3} \right) \Big|_0^{\sqrt{2}} \\
 &= \frac{1}{\sqrt{2}} \left(2\sqrt{2} - \frac{\sqrt{2}^3}{3} \right) \\
 &= 2 - \frac{2}{3} \\
 &= \frac{4}{3}.
 \end{aligned}$$

b. The intersection point of $y = f(x)$ and $y = r$ occurs, for positive x , when $2 - x^2 = \frac{4}{3}$, from which it follows that $x^2 = \frac{2}{3}$ and thus $x = \frac{\sqrt{6}}{3}$. A graph of f , $y = r$, and the intersection point is shown in the following figure.



c. Let A_1 be the area that lies below $y = f(x)$ and above $y = r$ on the interval $\left[0, \frac{\sqrt{6}}{3}\right]$ and A_2 the area of the region that lies below $y = r$ and above $y = f(x)$ on $\left[\frac{\sqrt{6}}{3}, \sqrt{2}\right]$ as illustrated in the figure below. Then

$$\begin{aligned}
 A_1 &= \int_0^{\sqrt{6}/3} (2 - x^2) - \frac{4}{3} \, dx \\
 &= \int_0^{\sqrt{6}/3} \frac{2}{3} - x^2 \, dx
 \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{2}{3}x - \frac{x^3}{3} \right) \bigg|_0^{\sqrt{6}/3} \\
&= \left(\frac{2}{3} \right) \left(\frac{\sqrt{6}}{3} \right) - \left(\frac{1}{3} \right) \left(\frac{\sqrt{6}}{3} \right)^3 \\
&= \frac{2\sqrt{6}}{9} - \frac{6\sqrt{6}}{81} \\
&= \frac{4\sqrt{6}}{27}.
\end{aligned}$$

and

$$\begin{aligned}
A_2 &= \int_{\sqrt{6}/3}^{\sqrt{2}} \frac{4}{3} - (2 - x^2) \, dx \\
&= \int_{\sqrt{6}/3}^{\sqrt{2}} x^2 - \frac{2}{3} \, dx \\
&= \left(\frac{x^3}{3} - \frac{2}{3}x \right) \bigg|_{\sqrt{6}/3}^{\sqrt{2}} \\
&= \left(\frac{2\sqrt{2}}{3} - \frac{2}{3}\sqrt{2} \right) - \left[\left(\frac{1}{3} \right) \left(\frac{\sqrt{6}}{3} \right)^3 - \left(\frac{2}{3} \right) \left(\frac{\sqrt{6}}{3} \right) \right] \\
&= \frac{4\sqrt{6}}{27}.
\end{aligned}$$

Hence $A_1 = A_2$ as desired.

- d. The answer is yes. Let AV is the average value of a function f on an interval $[a, b]$. By the definition of average value,

$$AV(b - a) = \int_a^b f(x) \, dx,$$

which can also be viewed as the area of the rectangle with base $(b - a)$ and height AV . Now suppose that the area of the region above the graph of f and below the line $y = AV$ on the interval $[a, b]$ is R_1 and the area of the region below the graph of f and above the line $y = AV$ on the interval $[a, b]$ is R_2 . The figure below illustrates that the value of the definite integral is the area of the rectangle plus the area above the rectangle minus the area below the rectangle, or

$$\int_a^b f(x) \, dx = AV(b - a) + R_2 - R_1.$$

Having already established that $AV(b - a) = \int_a^b f(x) \, dx$, it follows that $R_2 - R_1 = 0$. So the area of the region that lies below $y = f(x)$ and above $y = AV$ is equal to the amount of area that lies below $y = AV$ and above $y = f(x)$ on $[a, b]$.

6.2 Using Definite Integrals to Find Volume

Motivating Questions

- How can we use a definite integral to find the volume of a three-dimensional solid of revolution that results from revolving a two-dimensional region about a particular axis?
 - In what circumstances do we integrate with respect to y instead of integrating with respect to x ?
 - What adjustments do we need to make if we revolve about a line other than the x - or y -axis?
-

Just as we can use definite integrals to add the areas of rectangular slices to find the exact area that lies between two curves, we can also use integrals to find the volume of regions whose cross-sections have a particular shape.

In particular, we can determine the volume of solids whose cross-sections are all thin cylinders (or washers) by adding up the volumes of these individual slices. We first consider a familiar shape in Example 6.2.1: a circular cone.

Example 6.2.1. Consider a circular cone of radius 3 and height 5, which we view horizontally as pictured in Figure 6.2.2. Our goal in this example is to use a definite integral to determine the volume of the cone.

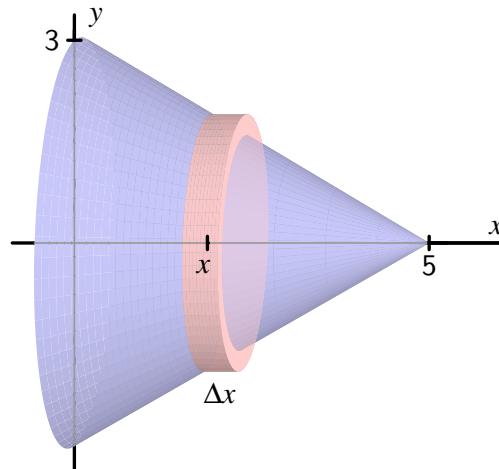


Figure 6.2.2: The circular cone described in Example 6.2.1

- a. Find a formula for the linear function $y = f(x)$ that is pictured in Figure 6.2.2.
- b. For the representative slice of thickness Δx that is located horizontally at a location x (somewhere between $x = 0$ and $x = 5$), what is the radius of the representative slice? Note that the radius depends on the value of x .
- c. What is the volume of the representative slice you found in (b)?
- d. What definite integral will sum the volumes of the thin slices across the full horizontal span of the cone? What is the exact value of this definite integral?
- e. Compare the result of your work in (d) to the volume of the cone that comes from using the formula $V_{\text{cone}} = \frac{1}{3}\pi r^2 h$.

6.2.1 The Volume of a Solid of Revolution

A solid of revolution is a three dimensional solid that can be generated by revolving one or more curves around a fixed axis. For example, the circular cone in Figure 6.2.2 is the solid of revolution generated by revolving the portion of the line $y = 3 - \frac{3}{5}x$ from $x = 0$ to $x = 5$ about the x -axis. Notice that if we slice a solid of revolution perpendicular to the axis of revolution, the resulting cross-section is a circle.

We first consider solids whose slices are thin cylinders. Recall that the volume of a cylinder is given by $V = \pi r^2 h$.

Example 6.2.3. Find the volume of the solid of revolution generated when the region R bounded by $y = 4 - x^2$ and the x -axis is revolved about the x -axis.

Solution. First, we observe that $y = 4 - x^2$ intersects the x -axis at the points $(-2, 0)$ and $(2, 0)$. When we revolve the region R about the x -axis, we get the three-dimensional solid pictured in Figure 6.2.4.

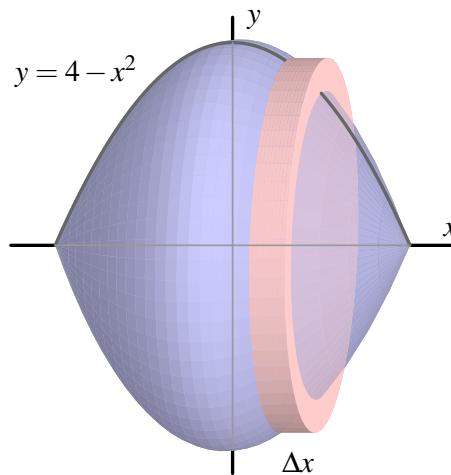


Figure 6.2.4: The solid of revolution in Example 6.2.3.

We slice the solid into vertical slices of thickness Δx between $x = -2$ and $x = 2$. A representative slice is a cylinder of height Δx and radius $4 - x^2$. Hence, the volume of the slice is

$$V_{\text{slice}} = \pi(4 - x^2)^2 \Delta x.$$

Using a definite integral to sum the volumes of the representative slices, it follows that

$$V = \int_{-2}^2 \pi(4 - x^2)^2 dx.$$

It is straightforward to evaluate the integral and find that the volume is $V = \frac{512}{15}\pi$.

For a solid such as the one in Example 6.2.3, where each slice is a cylindrical disk, we first find the volume of a typical slice (noting particularly how this volume depends on x), and then integrate over the range of x -values that bound the solid. Often, we will be content with simply finding the integral that represents the volume; if we desire a numeric value for the integral, we typically use a calculator or computer algebra system to find that value.

This method for finding the volume of a solid of revolution is often called the *disk method*.

The Disk Method

If $y = r(x)$ is a nonnegative continuous function on $[a, b]$, then the volume of the solid of revolution generated by revolving the curve about the x -axis over this interval is given by

$$V = \int_a^b \pi r(x)^2 dx.$$

A different type of solid can emerge when two curves are involved, as we see in the following example.

Example 6.2.5. Find the volume of the solid of revolution generated when the finite region R that lies between $y = 4 - x^2$ and $y = x + 2$ is revolved about the x -axis.

Solution. First, we must determine where the curves $y = 4 - x^2$ and $y = x + 2$ intersect. Substituting the expression for y from the second equation into the first equation, we find that $x + 2 = 4 - x^2$. Rearranging, it follows that

$$x^2 + x - 2 = 0,$$

and the solutions to this equation are $x = -2$ and $x = 1$. The curves therefore cross at $(-2, 0)$ and $(1, 1)$.

When we revolve the region R about the x -axis, we get the three-dimensional solid pictured at left in Figure 6.2.6.

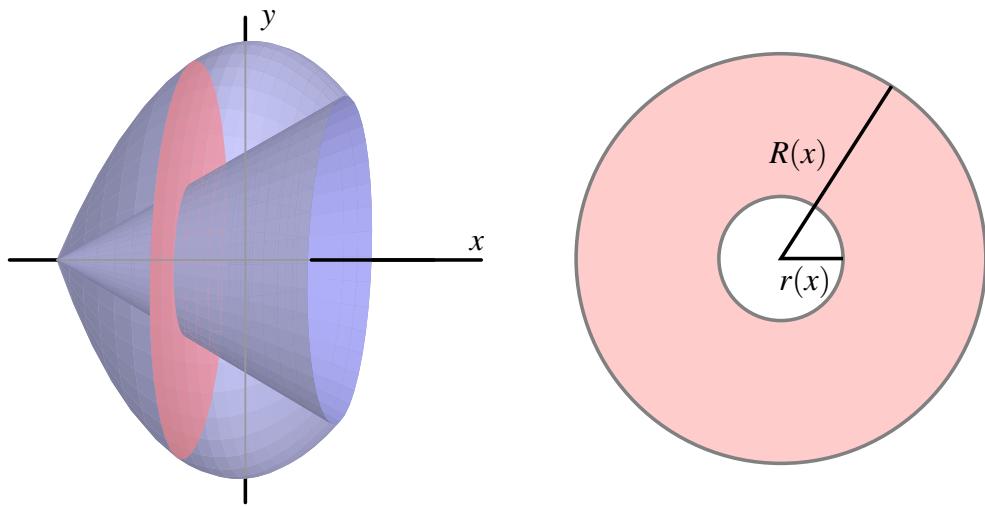


Figure 6.2.6: At left, the solid of revolution in Example 6.2.5. At right, a typical slice with inner radius $r(x)$ and outer radius $R(x)$.

Immediately we see a major difference between the solid in this example and the one in Example 6.2.3: here, the three-dimensional solid of revolution isn't "solid" because it has open space in its center along the axis of revolution. If we slice the solid perpendicular to the axis of revolution, we observe that the resulting slice is not a solid disk, but rather a *washer*, as pictured at right in Figure 6.2.6. At a given location x between $x = -2$ and $x = 1$, the small radius $r(x)$ of the inner circle is determined by the curve $y = x + 2$, so $r(x) = x + 2$. Similarly, the big radius $R(x)$ comes from the function $y = 4 - x^2$, and thus $R(x) = 4 - x^2$.

To find the volume of a representative slice, we compute the volume of the outer disk and subtract the volume of the inner disk. Since

$$\pi R(x)^2 \Delta x - \pi r(x)^2 \Delta x = \pi [R(x)^2 - r(x)^2] \Delta x,$$

it follows that the volume of a typical slice is

$$V_{\text{slice}} = \pi [(4 - x^2)^2 - (x + 2)^2] \Delta x.$$

Using a definite integral to sum the volumes of the respective slices across the integral, we find that

$$V = \int_{-2}^1 \pi [(4 - x^2)^2 - (x + 2)^2] dx.$$

Evaluating the integral, we find that the volume of the solid of revolution is $V = \frac{108}{5} \pi$.

This method for finding the volume of a solid of revolution generated by two curves is often called the *washer method*.

The Washer Method

If $y = R(x)$ and $y = r(x)$ are nonnegative continuous functions on $[a, b]$ that satisfy $R(x) \geq r(x)$ for all x in $[a, b]$, then the volume of the solid of revolution generated by revolving the region between them about the x -axis over this interval is given by

$$V = \int_a^b \pi[R(x)^2 - r(x)^2] dx.$$

Example 6.2.7. In each of the following questions, draw a careful, labeled sketch of the region described, as well as the resulting solid that results from revolving the region about the stated axis. In addition, draw a representative slice and state the volume of that slice, along with a definite integral whose value is the volume of the entire solid. It is not necessary to evaluate the integrals you find.

- The region S bounded by the x -axis, the curve $y = \sqrt{x}$, and the line $x = 4$; revolve S about the x -axis.
- The region S bounded by the y -axis, the curve $y = \sqrt{x}$, and the line $y = 2$; revolve S about the x -axis.
- The finite region S bounded by the curves $y = \sqrt{x}$ and $y = x^3$; revolve S about the x -axis.
- The finite region S bounded by the curves $y = 2x^2 + 1$ and $y = x^2 + 4$; revolve S about the x -axis.
- The region S bounded by the y -axis, the curve $y = \sqrt{x}$, and the line $y = 2$; revolve S about the y -axis. How is this problem different from the one posed in part (b)?

Hint.

- Use slices perpendicular to the x -axis.
- Note that slices perpendicular to the x -axis will be washers.
- Find the points where the two curves intersect and draw a labeled graph.
- Think about how the solid looks like a donut; what is the outer radius?
- Try using slices perpendicular to the y -axis.

Answer.

- $V = \int_0^4 \pi(\sqrt{x})^2 dx = \int_0^4 \pi x dx = 8\pi.$
- $V = \int_0^4 \pi(4 - (\sqrt{x})^2) dx = \int_0^4 \pi(4 - x) dx = 8\pi.$
- $V = \int_0^1 \pi(x - x^6) dx = \frac{5}{14}\pi.$

d. $V = \int_{-\sqrt{3}}^{\sqrt{3}} \pi((x^2 + 4)^2 - (2x^2 + 1)^2) dx = \frac{136\sqrt{3}}{5}\pi.$

e. $V = \int_0^2 \pi y^4 dy = \frac{32}{5}\pi.$

Solution.

- a. A typical slice (perpendicular to the x -axis) has radius $r(x) = \sqrt{x}$, and thus the volume of such a slice is

$$V_{\text{slice}} = \pi(\sqrt{x})^2 \Delta x.$$

It follows by the disk method that the volume of the solid of revolution is

$$V = \int_0^4 \pi(\sqrt{x})^2 dx = \int_0^4 \pi x dx = 8\pi.$$

- b. A typical slice (perpendicular to the x -axis) is a washer with outer radius $R(x) = 2$ and inner radius $r(x) = \sqrt{x}$. The volume of such a slice is

$$V_{\text{slice}} = \pi(R(x)^2 - r(x)^2) \Delta x = \pi(2^2 - (\sqrt{x})^2) \Delta x.$$

It follows by the washer method that the volume of the solid of revolution is

$$V = \int_0^4 \pi(4 - (\sqrt{x})^2) dx = \int_0^4 \pi(4 - x) dx = 8\pi.$$

- c. The curves $y = \sqrt{x}$ and $y = x^3$ intersect at $(0, 0)$ and $(1, 1)$; revolving the region S about the x -axis and taking slices perpendicular to the x -axis, we see that a typical slice is a washer with outer radius $R(x) = \sqrt{x}$ and inner radius $r(x) = x^3$. Hence the volume of a typical slice is

$$V_{\text{slice}} = \pi(R(x)^2 - r(x)^2) \Delta x = \pi((\sqrt{x})^2 - (x^3)^2) \Delta x = \pi(x - x^6) \Delta x.$$

By the washer method, the volume of the solid is

$$V = \int_0^1 \pi(x - x^6) dx = \frac{5}{14}\pi.$$

- d. The curves $y = 2x^2 + 1$ and $y = x^2 + 4$ intersect at the points $(\pm\sqrt{3}, 7)$; after we revolve the region S about the x -axis, cross-sections perpendicular to the axis of rotation are washers with outer radius $R(x) = x^2 + 4$ and inner radius $r(x) = 2x^2 + 1$. By the washer method, we see that

$$V = \int_{-\sqrt{3}}^{\sqrt{3}} \pi((x^2 + 4)^2 - (2x^2 + 1)^2) dx = \frac{136\sqrt{3}}{5}\pi.$$

- e. The curve $y = \sqrt{x}$ and the line $y = 2$ intersect at $(4, 2)$; when we revolve the region S about the y -axis, slices perpendicular to the y -axis are disks. Here we see that each disk's radius is a function of y and the disks have thickness Δy , so we will need to

integrate with respect to y and consider the radius function $r(y) = y^2$ (which comes from solving $y = \sqrt{x}$ for x in terms of y). Note how this is different from the problem in part (b) because of the axis we rotate about and how this forces us to integrate in terms of y . A typical slice has volume

$$V_{\text{slice}} = \pi(y^2)^2 \Delta y,$$

and thus by the disk method the total volume is

$$V = \int_0^2 \pi y^4 dy = \frac{32}{5}\pi.$$

6.2.2 Revolving about the y -axis

When we revolve a given region about the y -axis, the representative slices now have thickness Δy , which means that we must integrate with respect to y .

Example 6.2.8. Find the volume of the solid of revolution generated when the region R that lies between $y = \sqrt{x}$ and $y = x^4$ is revolved about the y -axis.

Solution. These two curves intersect when $x = 1$, hence at the point $(1, 1)$. When we revolve the region R about the y -axis, we get the three-dimensional solid pictured at left in Figure 6.2.9.

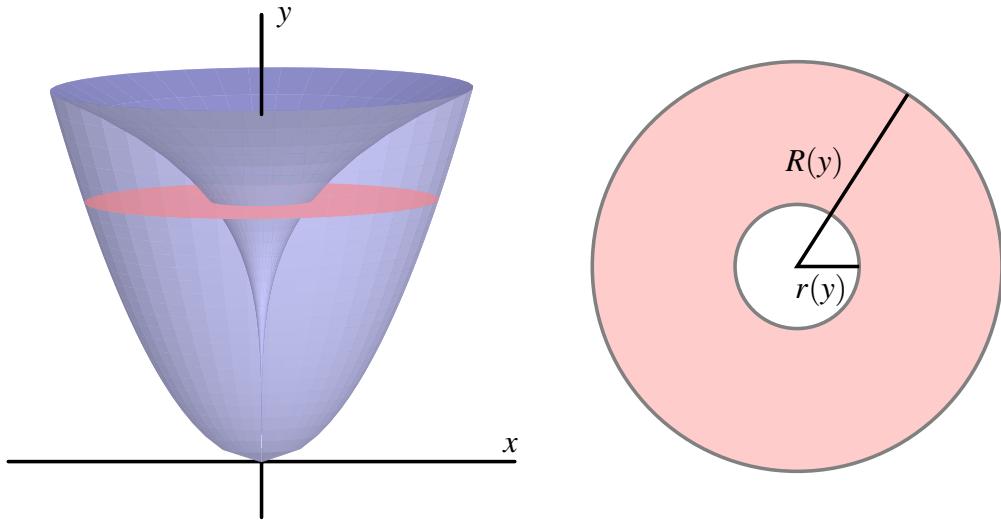


Figure 6.2.9: At left, the solid of revolution in Example 6.2.8. At right, a typical slice with inner radius $r(y)$ and outer radius $R(y)$.

Note that the slices are cylindrical washers only if taken perpendicular to the y -axis. We slice the solid horizontally, starting at $y = 0$ and proceeding up to $y = 1$. The thickness of a representative slice is Δy , so we must express the integrand in terms of y . The inner radius

is determined by the curve $y = \sqrt{x}$, so we solve for x and get $x = y^2 = r(y)$. In the same way, we solve the curve $y = x^4$ (which governs the outer radius) for x in terms of y , and hence $x = \sqrt[4]{y}$. Therefore, the volume of a typical slice is

$$V_{\text{slice}} = \pi[R(y)^2 - r(y)^2] = \pi[(\sqrt[4]{y})^2 - (y^2)^2]\Delta y.$$

We use a definite integral to sum the volumes of all the slices from $y = 0$ to $y = 1$. The total volume is

$$V = \int_{y=0}^{y=1} \pi [(\sqrt[4]{y})^2 - (y^2)^2] dy.$$

It is straightforward to evaluate the integral and find that $V = \frac{7}{15}\pi$.

Example 6.2.10. In each of the following questions, draw a careful, labeled sketch of the region described, as well as the resulting solid that results from revolving the region about the stated axis. In addition, draw a representative slice and state the volume of that slice, along with a definite integral whose value is the volume of the entire solid. It is not necessary to evaluate the integrals you find.

- The region S bounded by the y -axis, the curve $y = \sqrt{x}$, and the line $y = 2$; revolve S about the y -axis.
- The region S bounded by the x -axis, the curve $y = \sqrt{x}$, and the line $x = 4$; revolve S about the y -axis.
- The finite region S in the first quadrant bounded by the curves $y = 2x$ and $y = x^3$; revolve S about the x -axis.
- The finite region S in the first quadrant bounded by the curves $y = 2x$ and $y = x^3$; revolve S about the y -axis.
- The finite region S bounded by the curves $x = (y - 1)^2$ and $y = x - 1$; revolve S about the y -axis

Hint.

- Consider slices perpendicular to the y -axis.
- Since we revolve around the y -axis, think about slices of thickness Δy .
- Note that $y = x^3$ lies below $y = 2x$ on the interval $[0, \sqrt{2}]$.
- The curve $y = x^3$ can be equivalently expressed as $x = \sqrt[3]{y}$.
- Find the intersection points by writing the second curve as $x = y + 1$ and solving $(y - 1)^2 = y + 1$.

Answer.

- $V = \int_0^2 \pi y^4 \Delta y$.

- b. $V = \int_0^2 \pi(16 - y^4) dy.$
- c. $V = \int_0^{\sqrt{2}} \pi(4x^2 - x^6) dx.$
- d. $V = \int_0^{2\sqrt{2}} \pi(y^{2/3} - y^2/4) dy.$
- e. $V = \int_0^3 \pi((y+1)^2 - (y-1)^4) dy.$

Solution.

- a. Observe that horizontal slices with thickness Δy are all cylindrical disks, each with radius given by $r(y) = y^2$ since they are given by the curve $y = \sqrt{x}$, which can be equivalently expressed as $x = y^2$. The volume of each slice is thus

$$V_{\text{slice}} = \pi r(y)^2 \Delta y = \pi y^4 \Delta y.$$

The volume of the solid is thus given by the definite integral

$$V = \int_0^2 \pi y^4 \Delta y.$$

- b. Slicing perpendicular to the axis of rotation, we get slices of thickness Δy that are shaped like washers. The outer radius, $R(y)$, is given by $R(y) = 4$; the inner radius is given by the function $y = \sqrt{x}$, but expressed with x as a function of y , so $r(y) = y^2$. It follows that a representative washer slice has volume

$$V_{\text{slice}} = \pi(R(y)^2 - r(y)^2) \Delta y = \pi(16 - y^4) \Delta y.$$

Further, since $y = \sqrt{x}$ and $x = 4$ intersect at the point $(4, 2)$, we integrate from $y = 0$ to $y = 2$ and thus find that the total volume is

$$V = \int_0^2 \pi(16 - y^4) dy.$$

- c. The curves $y = x^3$ and $y = 2x$ intersect where $x^3 = 2x$, so where $x^3 - 2x = x(x^2 - 2) = 0$. The two nonnegative values of x that make this equation true are $x = 0$ and $x = \sqrt{2}$, and thus the points $(0, 0)$ and $(\sqrt{2}, 2\sqrt{2})$ form the boundaries of the region S that is being rotated about the x -axis. Since we rotate about the horizontal axis, we slice perpendicular to that axis with slices of thickness Δx . A representative slice is in the shape of a washer and has volume

$$V_{\text{slice}} = \pi(R(x)^2 - r(x)^2) \Delta x = \pi((2x)^2 - (x^3)^2) \Delta x,$$

and thus the total volume of the solid is

$$V = \int_0^{\sqrt{2}} \pi(4x^2 - x^6) dx.$$

- d. To use the same region S as in part (c), but now revolved about the y -axis, we shift to horizontal slices of thickness Δy , and thus need to express the curves as functions of y : the rightmost curve is $x = R(y) = \sqrt[3]{y}$, while the leftmost curve is $x = r(y) = \frac{1}{2}y$. A representative slice is a washer whose volume is

$$V_{\text{slice}} = \pi(R(y)^2 - r(y)^2)\Delta y = \pi((\sqrt[3]{y})^2 - (\frac{1}{2}y)^2)\Delta y.$$

It follows that the total volume of the solid is

$$V = \int_0^{2\sqrt{2}} \pi(y^{2/3} - y^2/4) dy.$$

- e. For this pair of curves, we see they intersect where $x = (y - 1)^2$ and $x = (y + 1)$ are both true, and hence where $(y - 1)^2 = y + 1$. Expanding, moving all terms to the left, and combining like terms, we see that $y^2 - 2y + 1 = y + 1$, so $y^2 - 3y = 0$. It follows that $y(y - 3) = 0$ and therefore $y = 3$ or $y = 0$. Plotting the region and revolving about the y -axis, we see that horizontal slices of thickness Δy are shaped like washers. The outer radius is given by $R(y) = y + 1$, while the inner radius is $r(y) = (y - 1)^2$. A representative slice thus has volume

$$V_{\text{slice}} = \pi(R(y)^2 - r(y)^2)\Delta y = \pi((y + 1)^2 - ((y - 1)^2)^2)\Delta y.$$

The volume of the solid of revolution is therefore

$$V = \int_0^3 \pi((y + 1)^2 - (y - 1)^4) dy.$$

6.2.3 Revolving about horizontal and vertical lines other than the coordinate axes

It is possible to revolve a region around any horizontal or vertical line. Doing so adjusts the radii of the cylinders or washers involved by a constant value. A careful, well-labeled plot of the solid of revolution will usually reveal how the different axis of revolution affects the definite integral.

Example 6.2.11. Find the volume of the solid of revolution generated when the finite region S that lies between $y = x^2$ and $y = x$ is revolved about the line $y = -1$.

Solution. Graphing the region between the two curves in the first quadrant between their points of intersection $((0, 0)$ and $(1, 1)$) and then revolving the region about the line $y = -1$, we see the solid shown in Figure 6.2.12. Each slice of the solid perpendicular to the axis of revolution is a washer, and the radii of each washer are governed by the curves $y = x^2$ and $y = x$. But we also see that there is one added change: the axis of revolution adds a fixed length to each radius. The inner radius of a typical slice, $r(x)$, is given by $r(x) = x^2 + 1$, while the outer radius is $R(x) = x + 1$.

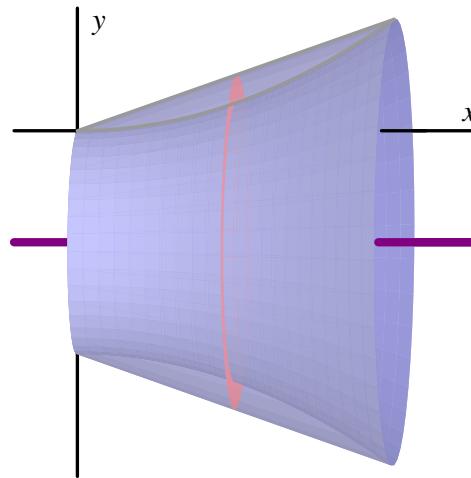


Figure 6.2.12: The solid of revolution described in Example 6.2.11.

Therefore, the volume of a typical slice is

$$V_{\text{slice}} = \pi[R(x)^2 - r(x)^2]\Delta x = \pi[(x+1)^2 - (x^2+1)^2]\Delta x.$$

Finally, we integrate to find the total volume, and

$$V = \int_0^1 \pi[(x+1)^2 - (x^2+1)^2] dx = \frac{7}{15}\pi.$$

Example 6.2.13. In each of the following questions, draw a careful, labeled sketch of the region described, as well as the resulting solid that results from revolving the region about the stated axis. In addition, draw a representative slice and state the volume of that slice, along with a definite integral whose value is the volume of the entire solid. It is not necessary to evaluate the integrals you find. For each prompt, use the finite region S in the first quadrant bounded by the curves $y = 2x$ and $y = x^3$.

- a. Revolve S about the line $y = -2$.
- b. Revolve S about the line $y = 4$.
- c. Revolve S about the line $x = -1$.
- d. Revolve S about the line $x = 5$.

Hint.

- a. Note that the two curves intersect in the first quadrant at $(0, 0)$ and $(\sqrt{2}, 2\sqrt{2})$.
- b. The outer radius is $R(x) = 4 - x^3$.
- c. Solve the equations governing the two curves for x in terms of y .
- d. The inner radius is $R(y) = 5 - y^{1/3}$.

Answer.

a. T

$$V = \int_0^{\sqrt{2}} \pi((2x+2)^2 - (x^3+2)^2) dx = \frac{4}{21}(21+8\sqrt{2})\pi \approx 19.336.$$

b.

$$V = \int_0^{\sqrt{2}} \pi((4-x^3)^2 - (4-2x)^2) dx = \left(8 - \frac{32\sqrt{2}}{21}\right)\pi \approx 18.3626.$$

c.

$$V = \int_0^{2\sqrt{2}} \pi((y^{1/3}+1)^2 - (\frac{1}{2}y+1)^2) dy = \frac{2}{15}(15+8\sqrt{2})\pi \approx 11.022.$$

d.

$$V = \int_0^{2\sqrt{2}} \pi((5-\frac{1}{2}y)^2 - (5-y^{1/3})^2) dy = \frac{2}{15}(75-8\sqrt{2})\pi \approx 26.677.$$

Solution.

- a. The two curves intersect in the first quadrant at $(0,0)$ and $(\sqrt{2}, 2\sqrt{2})$. Revolving the region S about the line $y = -2$, we see that a typical slice is a washer with thickness Δx , outer radius $R(x) = 2x+2$, and inner radius $r(x) = x^3+2$. It follows by the washer method that

$$V = \int_0^{\sqrt{2}} \pi((2x+2)^2 - (x^3+2)^2) dx = \frac{4}{21}(21+8\sqrt{2})\pi \approx 19.336.$$

- b. In this situation, revolving about $y = 4$, we see that the outer radius of a typical washer is $R(x) = 4 - x^3$, while the inner radius is $r(x) = 4 - 2x$, and thus the volume is

$$V = \int_0^{\sqrt{2}} \pi((4-x^3)^2 - (4-2x)^2) dx = \left(8 - \frac{32\sqrt{2}}{21}\right)\pi \approx 18.3626.$$

- c. Because we are revolving about $x = -1$, when we slice perpendicular to this we get washers of thickness Δy , and thus we need to write the two curves as functions of y in order to determine the radii of the slices as functions of y , we note that the line can be expressed as $x = \frac{1}{2}y$ and the cubic curve can be rewritten in the form $x = y^{1/3}$. The outer radius of a typical slice is thus $R(y) = y^{1/3}+1$ and the inner radius is $r(y) = \frac{1}{2}y+1$. It follows that the volume of the solid is

$$V = \int_0^{2\sqrt{2}} \pi((y^{1/3}+1)^2 - (\frac{1}{2}y+1)^2) dy = \frac{2}{15}(15+8\sqrt{2})\pi \approx 11.022.$$

- d. Here, the outer radius is $R(y) = 5 - \frac{1}{2}y$ and the inner radius is $r(y) = 5 - y^{1/3}$, so the total volume of the solid is

$$V = \int_0^{2\sqrt{2}} \pi((5-\frac{1}{2}y)^2 - (5-y^{1/3})^2) dy = \frac{2}{15}(75-8\sqrt{2})\pi \approx 26.677.$$

6.2.4 Summary

- We can use a definite integral to find the volume of a three-dimensional solid of revolution that results from revolving a two-dimensional region about a particular axis by taking slices perpendicular to the axis of revolution which will then be circular disks or washers.
- If we revolve about a vertical line and slice perpendicular to that line, then our slices are horizontal and of thickness Δy . This leads us to integrate with respect to y , as opposed to with respect to x when we slice a solid vertically.
- If we revolve about a line other than the x - or y -axis, we need to carefully account for the shift that occurs in the radius of a typical slice. Normally, this shift involves taking a sum or difference of the function along with the constant connected to the equation for the horizontal or vertical line; a well-labeled diagram is usually the best way to decide the new expression for the radius.

6.2.5 Exercises

1. (Solid of revolution from one function about the x -axis) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Solid of revolution from one function about the y -axis) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Solid of revolution from two functions about the x -axis) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Solid of revolution from two functions about a horizontal line) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Solid of revolution from two functions about a different horizontal line) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. (Solid of revolution from two functions about a vertical line) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
7. Consider the curve $f(x) = 3 \cos(\frac{x^3}{4})$ and the portion of its graph that lies in the first quadrant between the y -axis and the first positive value of x for which $f(x) = 0$. Let R denote the region bounded by this portion of f , the x -axis, and the y -axis.
 - a. Set up a definite integral whose value is the exact arc length of f that lies along the upper boundary of R . Use technology appropriately to evaluate the integral you find.
 - b. Set up a definite integral whose value is the exact area of R . Use technology appropriately to evaluate the integral you find.
 - c. Suppose that the region R is revolved around the x -axis. Set up a definite integral whose value is the exact volume of the solid of revolution that is generated. Use technology appropriately to evaluate the integral you find.

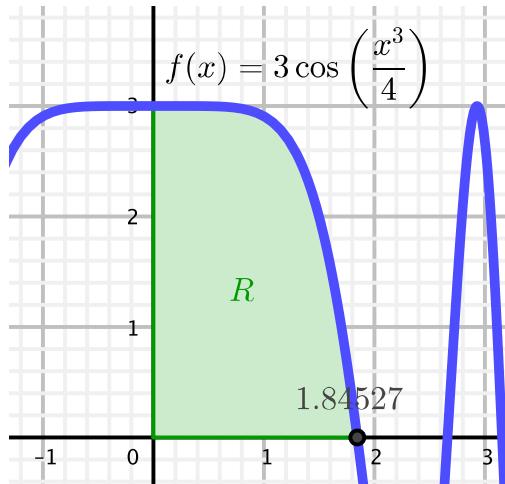
- d. Suppose instead that R is revolved around the y -axis. If possible, set up an integral expression whose value is the exact volume of the solid of revolution and evaluate the integral using appropriate technology. If not possible, explain why.

Answer.

- $L = \int_0^{1.842527} \sqrt{1 + \left(-3 \sin\left(\frac{x^3}{4}\right) \cdot \frac{3}{4}x^2\right)^2} dx \approx 4.10521.$
- $A = \int_0^{1.84527} 3 \cos\left(\frac{x^3}{4}\right) dx \approx 4.6623.$
- $V = \int_0^{1.84527} \pi \cdot 9 \cos^2\left(\frac{x^3}{4}\right) dx \approx 40.31965.$
- $V = \int_0^3 \pi (4 \arccos(\frac{y}{3}))^{2/3} dy \approx 23.29194.$

Solution.

- a. Plotting the function and using technology to find where $f(x)$ intersects the x -axis, we see that the region R as pictured in the following figure.



To apply the formula we developed for arc length, we need the derivative of f . By the chain rule, $f'(x) = -3 \sin\left(\frac{x^3}{4}\right) \cdot \frac{3}{4}x^2$, so the arc length of f that lies along the upper boundary of R is

$$L = \int_0^{1.842527} \sqrt{1 + \left(-3 \sin\left(\frac{x^3}{4}\right) \cdot \frac{3}{4}x^2\right)^2} dx.$$

Using technology to evaluate the integral, we find $L \approx 4.10521$.

- b. The area of R is

$$A = \int_0^{1.84527} 3 \cos\left(\frac{x^3}{4}\right) dx \approx 4.6623.$$

- c. Revolving R around the x -axis, we can use the disk method with slices perpendicular to the x -axis to find the volume of the resulting region. Noting that the radius of each such disk is $r(x) = 3 \cos(\frac{x^3}{4})$ and the width of each such disk is Δx , we know the volume of each slice is

$$V_{\text{slice}} \approx \pi r(x)^2 \Delta x = \pi \left(3 \cos\left(\frac{x^3}{4}\right)\right)^2 \Delta x$$

and thus the total volume is

$$V = \int_0^{1.84527} \pi \cdot 9 \cos^2 \left(\frac{x^3}{4} \right) dx \approx 40.31965$$

- d. If we instead revolve R around the y -axis, we need to use horizontal disks of thickness Δy in order to apply the disk method. To do that, we also need to express x as a function of y , rather than y as a function of x as given. Thus we take the given function $y = 3 \cos(\frac{x^3}{4})$ and solve the equation for x . First, we observe

$$\cos \left(\frac{x^3}{4} \right) = \frac{y}{3}$$

and thus using the inverse cosine,

$$\frac{x^3}{4} = \arccos \left(\frac{y}{3} \right)$$

and therefore $x^3 = 4 \arccos(\frac{y}{3})$ so that

$$x = \sqrt[3]{4 \arccos \left(\frac{y}{3} \right)}.$$

Now, a typical slice has radius $r(y) = \sqrt[3]{4 \arccos(\frac{y}{3})}$ with thickness Δy , and volume

$$V_{\text{slice}} \approx \pi \cdot \left(\sqrt[3]{4 \arccos \left(\frac{y}{3} \right)} \right)^2 \Delta y.$$

Integrating to add up the slices from $y = 0$ to $y = 3$, we have the volume of the solid of revolution is

$$V = \int_0^3 \pi \left(4 \arccos \left(\frac{y}{3} \right) \right)^{2/3} dy \approx 23.29194.$$

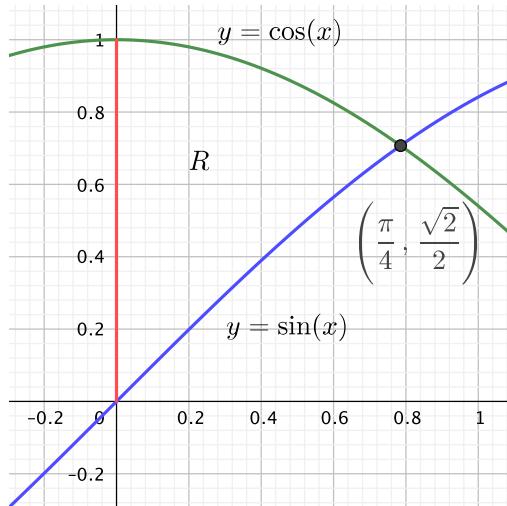
8. Consider the curves given by $y = \sin(x)$ and $y = \cos(x)$. For each of the following problems, you should include a sketch of the region/solid being considered, as well as a labeled representative slice.

- Sketch the region R bounded by the y -axis and the curves $y = \sin(x)$ and $y = \cos(x)$ up to the first positive value of x at which they intersect. What is the exact intersection point of the curves?
- Set up a definite integral whose value is the exact area of R .
- Set up a definite integral whose value is the exact volume of the solid of revolution generated by revolving R about the x -axis.
- Set up a definite integral whose value is the exact volume of the solid of revolution generated by revolving R about the y -axis.
- Set up a definite integral whose value is the exact volume of the solid of revolution generated by revolving R about the line $y = 2$.

- f. Set up a definite integral whose value is the exact volume of the solid of revolution generated by revolving R about the line $x = -1$.

Answer.

a.



b. $A = \int_0^{\frac{\pi}{4}} (\cos(x) - \sin(x)) dx$.

c. $V = \int_0^{\frac{\pi}{4}} \pi (\cos^2(x) - \sin^2(x)) dx$.

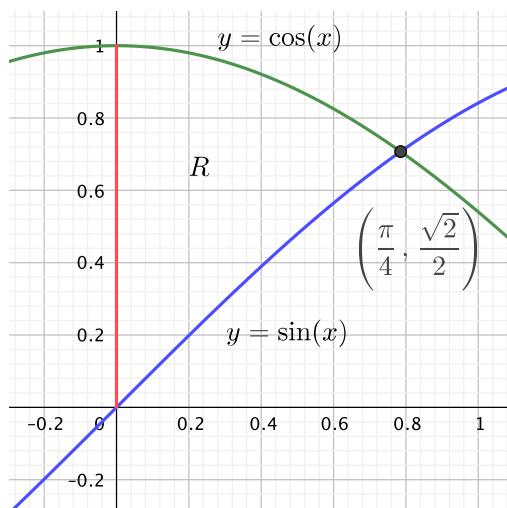
d. $V = \int_0^{\frac{\sqrt{2}}{2}} \pi \arcsin^2(y) dy + \int_{\frac{\sqrt{2}}{2}}^1 \pi \arccos^2(y) dy$

e. $V = \int_0^{\frac{\pi}{4}} \pi [(2 - \sin(x))^2 - (2 - \cos(x))^2] dx$.

f. $V = \int_0^{\frac{\sqrt{2}}{2}} \pi [(1 + \arcsin(y))^2 - 1^2] dy + \int_{\frac{\sqrt{2}}{2}}^1 \pi [(1 + \arccos(y))^2 - 1^2] dy$

Solution.

a.



See the sketch above. The exact intersection point of the curves is where $\cos(x) = \sin(x)$ in the first quadrant, and thus at the point $(\frac{\pi}{4}, \frac{\sqrt{2}}{2})$.

- b. The area of R is given by the definite integral

$$A = \int_0^{\frac{\pi}{4}} (\cos(x) - \sin(x)) dx.$$

- c. Revolving R about the x -axis and using slices perpendicular to the axis with thickness Δx , the outer radius of the resulting washers is $R(x) = \cos(x)$, while the inner radius of the washers is $r(x) = \sin(x)$. The volume of a typical slice is

$$V_{\text{slice}} \approx \pi(\cos^2(x) - \sin^2(x))\Delta x$$

and thus using an integral to sum the slices over the interval $x = 0$ to $x = \pi/4$, we find the volume of the solid is

$$V = \int_0^{\frac{\pi}{4}} \pi(\cos^2(x) - \sin^2(x)) dx.$$

- d. Revolving R about the y -axis, the resulting figure has solid cross sections that are disks perpendicular to the y -axis with thickness Δx , but the radius $R(y)$ that governs the radius of each disk changes at the height where $y = \frac{\sqrt{2}}{2}$ in the interval from $y = 0$ to $y = 1$. We also need to express the two curves that bound the region with x as a function of y . On the interval from $y = 0$ to $y = \frac{\sqrt{2}}{2}$, the sine function provides the boundary, which we write as $x = \arcsin(y)$; on the interval from $y = \frac{\sqrt{2}}{2}$ to $y = 1$, the cosine function provides the boundary, which we write as $x = \arccos(y)$. On the lower interval, a typical slice has volume

$$V_{\text{slice}} \approx \pi \arcsin^2(y) \Delta y.$$

Using similar reasoning for the upper interval and summing the slices using appropriate definite integrals, we find that

$$V = \int_0^{\frac{\sqrt{2}}{2}} \pi \arcsin^2(y) dy + \int_{\frac{\sqrt{2}}{2}}^1 \pi \arccos^2(y) dy$$

- e. Draw a picture. Revolving R about the line $y = 2$, we see that cross-sections perpendicular to the line $y = 2$ are washers with thickness Δx , outer radius $R(x) = 2 - \sin(x)$, and inner radius $r(x) = 2 - \cos(x)$, since $2 - \sin(x) \geq 2 - \cos(x)$ on the interval $[0, \frac{\pi}{4}]$. Thus, the volume of a typical slice is

$$V_{\text{slice}} \approx \pi[(2 - \sin(x))^2 - (2 - \cos(x))^2] \Delta x$$

and hence integrating to sum the total volume, we find

$$V = \int_0^{\frac{\pi}{4}} \pi[(2 - \sin(x))^2 - (2 - \cos(x))^2] dx.$$

- f. Revolving R about the line $x = -1$, the resulting figure has solid cross sections that are washers perpendicular to the y -axis with thickness Δy ; the inner radius is always $r(y) = 1$, but the outer radius $R(y)$ that governs the larger radius of each washer changes at the height where $y = \frac{\sqrt{2}}{2}$ in the interval from $y = 0$ to $y = 1$. As in (d), we express the two curves that bound the region with x as a function of y . On the interval from $y = 0$ to $y = \frac{\sqrt{2}}{2}$, $R(y) = 1 + \arcsin(y)$; on the interval from $y = \frac{\sqrt{2}}{2}$ to $y = 1$, $R(y) = 1 + \arccos(y)$. On the lower interval, a typical slice has volume

$$V_{\text{slice}} \approx \pi[(1 + \arcsin(y))^2 - 1^2]\Delta y.$$

Using similar reasoning for the upper interval and summing the slices using appropriate definite integrals, we find that

$$V = \int_0^{\frac{\sqrt{2}}{2}} \pi[(1 + \arcsin(y))^2 - 1^2] dy + \int_{\frac{\sqrt{2}}{2}}^1 \pi[(1 + \arccos(y))^2 - 1^2] dy$$

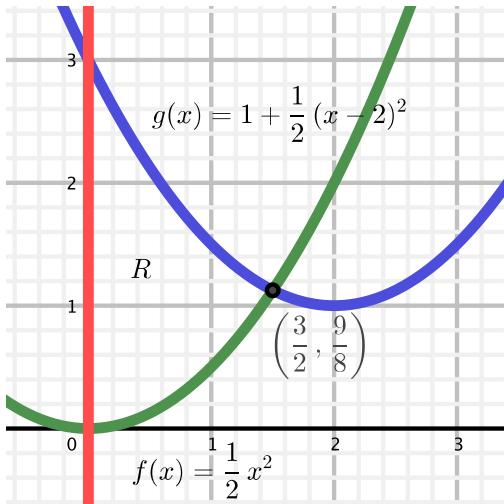
9. Consider the finite region R that is bounded by the curves $y = 1 + \frac{1}{2}(x - 2)^2$, $y = \frac{1}{2}x^2$, and $x = 0$.
- Determine a definite integral whose value is the area of the region enclosed by the two curves.
 - Find an expression involving one or more definite integrals whose value is the volume of the solid of revolution generated by revolving the region R about the line $y = -1$.
 - Determine an expression involving one or more definite integrals whose value is the volume of the solid of revolution generated by revolving the region R about the y -axis.
 - Find an expression involving one or more definite integrals whose value is the perimeter of the region R .

Answer.

- $A = \int_0^{1.5} 1 + \frac{1}{2}(x - 2)^2 - \frac{1}{2}x^2 dx = 2.25.$
- $V = \int_0^{1.5} \pi \left[\left(2 + \frac{1}{2}(x - 2)^2\right)^2 - \left(1 + \frac{1}{2}x^2\right)^2 \right] dx = \frac{315}{32}\pi$
- $V = \int_0^{1.125} \pi (\sqrt{2y})^2 dy + \int_{1.125}^3 \pi (2 - \sqrt{2(y-1)})^2 dy \approx 7.06858347.$
- $P = 3 + \int_0^{1.5} \sqrt{1 + (x - 2)^2} + \sqrt{1 + x^2} dx \approx 7.387234642.$

Solution.

a.



A picture of the region is shown in the figure above. The point of intersection occurs when

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

so $2 + (x - 2)^2 = x^2$, which implies $2 + (x^2 - 4x + 4) = x^2$ and thus $4x = 6$ so $x = \frac{3}{2}$. Since $g(x) = 1 + \frac{1}{2}(x - 2)^2 \geq \frac{1}{2}x^2 = f(x)$ on the interval $[0, \frac{3}{2}]$, the area of the region between the two curves is

$$\begin{aligned} A &= \int_0^{1.5} 1 + \frac{1}{2}(x - 2)^2 - \frac{1}{2}x^2 \, dx \\ &= \int_0^{1.5} 1 + \frac{1}{2}(x^2 - 4x + 4) - \frac{1}{2}x^2 \, dx \\ &= \int_0^{1.5} 3 - 2x \, dx \\ &= (3x - x^2) \Big|_0^{1.5} \\ &= 4.5 - 2.25 \\ &= 2.25. \end{aligned}$$

- b. (You should draw a picture of the solid, based on the figure above, with the region R revolved about $y = -1$.) We slice vertically so that our cross sections are washers with thickness Δx . Because we are revolving about $y = -1$, the outer radius is $R(x) = 2 + \frac{1}{2}(x - 2)^2$, while the inner radius is $1 + \frac{1}{2}x^2$. Hence the volume of a typical slice, which has thickness Δx is given by

$$V_{\text{slice}} \approx \pi (R(x)^2 - r(x)^2) \Delta x = \pi \left[\left(2 + \frac{1}{2}(x - 2)^2\right)^2 - \left(1 + \frac{1}{2}x^2\right)^2 \right] \Delta x.$$

Hence, the volume of the solid obtained by revolving the region R around the line $y = -1$ is given by

$$V = \int_0^{1.5} \pi \left[\left(2 + \frac{1}{2}(x - 2)^2\right)^2 - \left(1 + \frac{1}{2}x^2\right)^2 \right] dx = \frac{315}{32}\pi$$

- c. (Draw a picture of the solid, based on revolving the region R about the y -axis.) We slice horizontally and thus the resulting cross sections are cylinders with thickness Δy . The radius of the disks changes as y passes through $f\left(\frac{3}{2}\right) = \frac{9}{8} = 1.125$, so we need two different integrals to find the total volume. Since we are slicing horizontally, we will integrate with respect to y , so we need to solve both of our equations for x as functions of y . Observe that $y = 1 + \frac{1}{2}(x - 2)^2$ implies $2(y - 1) = (x - 2)^2$, so $x - 2 = \pm\sqrt{2(y - 1)}$ and thus $x = 2 \pm \sqrt{2(y - 1)}$, and similarly $y = \frac{1}{2}x^2$ implies $x = \pm\sqrt{2y}$. The portion of the curve defined by $y = 1 + \frac{1}{2}(x - 2)^2$ that determines one of the boundaries is the left hand portion of the parabola, so $x = 2 - \sqrt{2(y - 1)}$, while the portion of the curve $y = \frac{1}{2}x^2$ that determines the other boundary is the right hand portion of the parabola, so $x = \sqrt{2y}$. These two curves define the radius of the disk (one for $y < 1.125$, the other for $y > 1.125$) and the lowest point on the region is $(0, 0)$ while the highest is $(0, 3)$. So a sum of integrals that represents the volume of the solid obtained by revolving the region R around the y -axis is

$$V = \int_0^{1.125} \pi (\sqrt{2y})^2 dy + \int_{1.125}^3 \pi (2 - \sqrt{2(y - 1)})^2 dy \approx 7.06858347$$

- d. Recall that the length of a curve defined by a function f on an interval $[a, b]$ is given by

$$\int_a^b \sqrt{1 + (f'(x))^2} dx.$$

To find the perimeter of the region R we calculate the length of the curves defined by $y = 1 + \frac{1}{2}(x - 2)^2$ (for which $y' = x - 2$) and $y = \frac{1}{2}x^2$ (for which $y' = x$) on the interval $[0, 1.5]$ and the length of the segment along $x = 0$ as y goes from 0 to 3. This last length is 3, so the perimeter of the region R is given by

$$P = 3 + \int_0^{1.5} \sqrt{1 + (x - 2)^2} + \sqrt{1 + x^2} dx \approx 7.387234642.$$

6.3 Density, Mass, and Center of Mass

Motivating Questions

- How are mass, density, and volume related?
- How is the mass of an object with varying density computed?
- What is the center of mass of an object, and how are definite integrals used to compute it?

Studying the units of the integrand and variable of integration helps us understand the meaning of a definite integral. For instance, if $v(t)$ is the velocity of an object moving along an axis, measured in feet per second, and t measures time in seconds, then both the definite integral and its Riemann sum approximation,

$$\int_a^b v(t) dt \approx \sum_{i=1}^n v(t_i) \Delta t,$$

have units given by the product of the units of $v(t)$ and t :

$$(\text{feet/sec}) \cdot (\text{sec}) = \text{feet}.$$

Thus, $\int_a^b v(t) dt$ measures the total change in position of the moving object in feet.

Unit analysis will be particularly helpful to us in what follows.

Example 6.3.1. In each of the following scenarios, we consider the distribution of a quantity along an axis.

- a. Suppose that the function $c(x) = 200 + 100e^{-0.1x}$ models the density of traffic on a straight road, measured in cars per mile, where x is number of miles east of a major interchange, and consider the definite integral $\int_0^2 (200 + 100e^{-0.1x}) dx$.

- What are the units of the product $c(x) \cdot \Delta x$?
- What are the units of the definite integral and its Riemann sum approximation given by

$$\int_0^2 c(x) dx \approx \sum_{i=1}^n c(x_i) \Delta x?$$

- Evaluate the definite integral $\int_0^2 c(x) dx = \int_0^2 (200 + 100e^{-0.1x}) dx$ and write one sentence to explain the meaning of the value you find.

- b. On a 6 foot long shelf filled with books, the function B models the distribution of the weight of the books, in pounds per inch, where x is the number of inches from the left end of the bookshelf. Let $B(x)$ be given by the rule $B(x) = 0.5 + \frac{1}{(x+1)^2}$.

- What are the units of the product $B(x) \cdot \Delta x$?
- What are the units of the definite integral and its Riemann sum approximation given by

$$\int_{12}^{36} B(x) dx \approx \sum_{i=1}^n B(x_i) \Delta x?$$

- Evaluate the definite integral $\int_0^{72} B(x) dx = \int_0^{72} \left(0.5 + \frac{1}{(x+1)^2}\right) dx$ and write one sentence to explain the meaning of the value you find.

6.3.1 Density

The *mass* of a quantity, typically measured in metric units such as grams or kilograms, is a measure of the amount of the quantity. In a corresponding way, the *density* of an object measures the distribution of mass per unit volume. For instance, if a brick has mass 3 kg and volume 0.002 m³, then the density of the brick is

$$\frac{3\text{kg}}{0.002\text{m}^3} = 1500 \frac{\text{kg}}{\text{m}^3}.$$

As another example, the mass density of water is 1000 kg/m³. Each of these relationships demonstrate the following general principle.

For an object of constant density d , with mass m and volume V ,

$$d = \frac{m}{V}, \text{ or } m = d \cdot V.$$

But what happens when the density is not constant?

The formula $m = d \cdot V$ is reminiscent of two other equations that we have used in our work: for a body moving in a fixed direction, distance = rate \cdot time, and, for a rectangle, its area is given by $A = l \cdot w$. These formulas hold when the principal quantities involved, such as the rate the body moves and the height of the rectangle, are *constant*. When these quantities are not constant, we have turned to the definite integral for assistance. By working with small slices on which the quantity of interest (such as velocity) is approximately constant, we can use a definite integral to add up the values on the pieces.

For example, if we have a nonnegative velocity function that is not constant, over a short time interval Δt we know that the distance traveled is approximately $v(t)\Delta t$, since $v(t)$ is almost constant on a small interval. Similarly, if we are thinking about the area under a nonnegative function f whose value is changing, on a short interval Δx the area under the curve is approximately the area of the rectangle whose height is $f(x)$ and whose width is Δx : $f(x)\Delta x$. Both of these principles are represented visually in Figure 6.3.2.

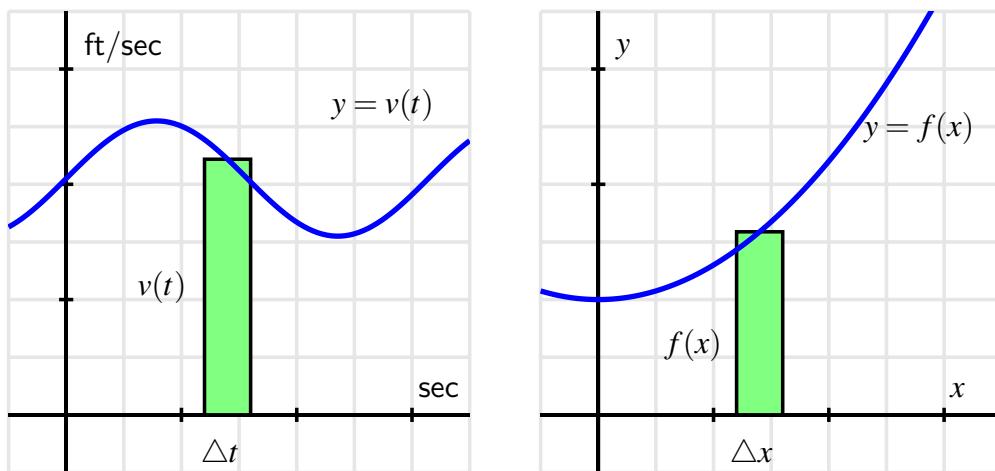


Figure 6.3.2: At left, estimating a small amount of distance traveled, $v(t)\Delta t$, and at right, a small amount of area under the curve, $f(x)\Delta x$.

In a similar way, if the density of some object is not constant, we can use a definite integral to compute the overall mass of the object. We will focus on problems where the density varies in only one dimension, say along a single axis.

Let's consider a thin bar of length b whose left end is at the origin, where $x = 0$, and assume that the bar has constant cross-sectional area of 1 cm^2 . We let $\rho(x)$ represent the mass density function of the bar, measured in grams per cubic centimeter. That is, given a location x , $\rho(x)$ tells us approximately how much mass will be found in a one-centimeter wide slice of the bar at x .

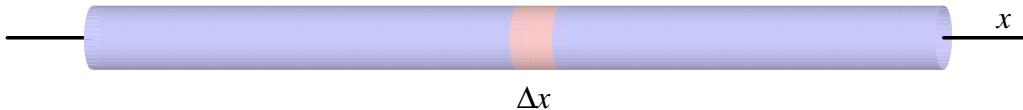


Figure 6.3.3: A thin bar of constant cross-sectional area 1 cm^2 with density function $\rho(x) \text{ g/cm}^3$.

The volume of a thin slice of the bar of width Δx , as pictured in Figure 6.3.3, is the cross-sectional area times Δx . Since the cross-sections each have constant area 1 cm^2 , it follows that the volume of the slice is $1\Delta x \text{ cm}^3$. And because mass is the product of density and volume, we see that the mass of this slice is approximately

$$\text{mass}_{\text{slice}} \approx \rho(x) \frac{\text{g}}{\text{cm}^3} \cdot 1\Delta x \text{ cm}^3 = \rho(x) \cdot \Delta x \text{ g.}$$

The corresponding Riemann sum (and the integral that it approximates),

$$\sum_{i=1}^n \rho(x_i) \Delta x \approx \int_0^b \rho(x) dx,$$

therefore measure the mass of the bar between 0 and b . (The Riemann sum is an approximation, while the integral will be the exact mass.)

For objects whose cross-sectional area is constant and whose mass is distributed relative to horizontal location, x , it makes sense to think of the density function $\rho(x)$ with units “mass per unit length,” such as g/cm. Thus, when we compute $\rho(x) \cdot \Delta x$ on a small slice Δx , the resulting units are g/cm · cm = g, which thus measures the mass of the slice. The general principle follows.

For an object of constant cross-sectional area whose mass is distributed along a single axis according to the function $\rho(x)$ (whose units are units of mass per unit of length), the total mass, M , of the object between $x = a$ and $x = b$ is given by

$$M = \int_a^b \rho(x) dx.$$

Example 6.3.4. Consider the following situations in which mass is distributed in a non-constant manner.

- a. Suppose that a thin rod with constant cross-sectional area of 1 cm^2 has its mass distributed according to the density function $\rho(x) = 2e^{-0.2x}$, where x is the distance in cm from the left end of the rod, and the units of $\rho(x)$ are g/cm. If the rod is 10 cm long, determine the exact mass of the rod.
- b. Consider the cone that has a base of radius 4 m and a height of 5 m. Picture the cone lying horizontally with the center of its base at the origin and think of the cone as a solid of revolution.
 - i. Write and evaluate a definite integral whose value is the volume of the cone.
 - ii. Next, suppose that the cone has uniform density of 800 kg/m^3 . What is the mass of the solid cone?
 - iii. Now suppose that the cone’s density is not uniform, but rather that the cone is most dense at its base. In particular, assume that the density of the cone is uniform across cross sections parallel to its base, but that in each such cross section that is a distance x units from the origin, the density of the cross section is given by the function $\rho(x) = 400 + \frac{200}{1+x^2}$, measured in kg/m³. Determine and evaluate a definite integral whose value is the mass of this cone of non-uniform density. Do so by first thinking about the mass of a given slice of the cone x units away from the base; remember that in such a slice, the density will be *essentially constant*.
- c. Let a thin rod of constant cross-sectional area 1 cm^2 and length 12 cm have its mass be distributed according to the density function $\rho(x) = \frac{1}{25}(x - 15)^2$, measured in g/cm. Find the exact location z at which to cut the bar so that the two pieces will each have identical mass.

Hint.

- Remember that $M = \int_a^b \rho(x) dx$.
- Think about slices: what's the volume of a slice? what's the mass of a slice?
- Consider $\int_0^b \rho(x) dx$.

Answer.

- $M = 0.1 - 0.1e^{-4} \approx 0.0981684$ grams.
- $V = \int_0^5 \pi(4 - \frac{4}{5}x)^2 dx = \frac{80\pi}{3} \approx 83.7758m^3$.
 - $M = \int_0^5 (400 + \frac{200}{1+x^2}) \cdot \pi(4 - \frac{4}{5}x)^2 dx = \frac{32}{5}\pi(25015 - 5\ln(17576) + 72\tan^{-1}(5)) \approx 33597.4664\text{kg}$
- $b \approx 3.0652$.

Solution.

- Since the mass, M , is given by $M = \int_a^b \rho(x) dx$, it follows that

$$M = \int_0^{10} 2e^{-0.2x} = \frac{2}{-0.2}e^{-0.2x} \Big|_0^{20} = -0.1e^{-4} + 0.1e^0,$$

and hence $M = 0.1 - 0.1e^{-4} \approx 0.0981684$ grams.

- Consider the cone that has a base of radius 4 m and a height of 5 m.
 - With the cone having its base centered at the origin and being a solid of revolution about the x -axis, the cone is generated by the line $y = 4 - \frac{4}{5}x$. Using the disk method, it follows that the cone's volume is

$$V = \int_0^5 \pi(4 - \frac{4}{5}x)^2 dx = \frac{80\pi}{3} \approx 83.7758m^3.$$

- If the cone has uniform density 800 kg/m^3 , we know that the cone's mass is the product of its density and volume, and thus

$$M = \rho \cdot V = 800 \cdot \frac{80\pi}{3} = \frac{6400\pi}{3} \approx 67020.6433\text{kg}.$$

- If the cone has non-uniform density, but is uniform across cross sections parallel to its base such that in each cross section a distance x units from the origin has its density given by $\rho(x) = 400 + \frac{200}{1+x^2} \text{ kg/m}^3$, we naturally consider slices of the cone that are perpendicular to the x -axis. For each such slice, the radius is $r(x) = 4 - \frac{4}{5}x$, which makes the volume of the slice

$$V_{\text{slice}} \approx \pi(4 - \frac{4}{5}x)^2 \Delta x.$$

Since the density is constant on the slice with value $\rho(x) = 400 + \frac{200}{1+x^2}$, it follows that the mass of the slice is

$$M_{\text{slice}} \approx (400 + \frac{200}{1+x^2}) \cdot \pi(4 - \frac{4}{5}x)^2 \Delta x.$$

Using a definite integral to let $\Delta x \rightarrow 0$ and add up the values of all of the slices, we thus find the mass of the cone with this density distribution to be

$$\begin{aligned} M &= \int_0^5 (400 + \frac{200}{1+x^2}) \cdot \pi(4 - \frac{4}{5}x)^2 dx \\ &= \frac{32}{5} \pi (25015 - 5 \ln(17576) + 72 \tan^{-1}(5)) \approx 33597.4664 \text{ kg} \\ &\approx 33597.4664 \text{ kg}. \end{aligned}$$

- c. The total mass of the bar is given by $M = \int_0^{12} \frac{1}{25}(x - 15)^2 dx = \frac{1116}{25}$. To find where to cut the bar in two so that the two pieces have equal mass, we have to determine the value of b such that

$$\int_0^b \frac{1}{25}(x - 15)^2 dx = \frac{1}{2} \cdot \frac{1116}{25}.$$

Since $\int_0^b \frac{1}{25}(x - 15)^2 dx = \frac{1}{75}((b - 15)^3 - (-15)^3)$, we need to have b satisfy the equation

$$\frac{1}{75}((b - 15)^3 - (-15)^3) = \frac{1}{2} \cdot \frac{1116}{25}.$$

Thus, $(b - 15)^3 - (-15)^3 = 75 \cdot \frac{558}{25} = 3 \cdot 558 = 1674$, so $(b - 15)^3 = 1675 - 3375 = -1700$. It follows that $b - 15 = \sqrt[3]{-1700} \approx -11.9348$, so $b \approx 3.0652$.

6.3.2 Weighted Averages

The concept of an average is a natural one, and one that we have used repeatedly as part of our understanding of the meaning of the definite integral. If we have n values a_1, a_2, \dots, a_n , we know that their average is given by

$$\frac{a_1 + a_2 + \dots + a_n}{n},$$

and for a quantity being measured by a function f on an interval $[a, b]$, the average value of the quantity on $[a, b]$ is

$$\frac{1}{b-a} \int_a^b f(x) dx.$$

As we continue to think about problems involving the distribution of mass, it is natural to consider the idea of a *weighted* average, where certain quantities involved are counted more in the average.

A common use of weighted averages is in the computation of a student's GPA, where grades are weighted according to credit hours. Let's consider the scenario in Table 6.3.5.

class	grade	grade points	credits
chemistry	B+	3.3	5
calculus	A-	3.7	4
history	B-	2.7	3
psychology	B-	2.7	3

Table 6.3.5: A college student's semester grades.

If all of the classes were of the same weight (i.e., the same number of credits), the student's GPA would simply be calculated by taking the average

$$\frac{3.3 + 3.7 + 2.7 + 2.7}{4} = 3.1.$$

But since the chemistry and calculus courses have higher weights (of 5 and 4 credits respectively), we actually compute the GPA according to the weighted average

$$\frac{3.3 \cdot 5 + 3.7 \cdot 4 + 2.7 \cdot 3 + 2.7 \cdot 3}{5 + 4 + 3 + 3} = 3.1\bar{6}.$$

The weighted average reflects the fact that chemistry and calculus, as courses with higher credits, have a greater impact on the students' grade point average. Note particularly that in the weighted average, each grade gets multiplied by its weight, and we divide by the sum of the weights.

In the following example, we explore further how weighted averages can be used to find the balancing point of a physical system.

Example 6.3.6. For quantities of equal weight, such as two children on a teeter-totter, the balancing point is found by taking the average of their locations. When the weights of the quantities differ, we use a weighted average of their respective locations to find the balancing point.

- Suppose that a shelf is 6 feet long, with its left end situated at $x = 0$. If one book of weight 1 lb is placed at $x_1 = 0$, and another book of weight 1 lb is placed at $x_2 = 6$, what is the location of \bar{x} , the point at which the shelf would (theoretically) balance on a fulcrum?
- Now, say that we place four books on the shelf, each weighing 1 lb: at $x_1 = 0$, at $x_2 = 2$, at $x_3 = 4$, and at $x_4 = 6$. Find \bar{x} , the balancing point of the shelf.
- How does \bar{x} change if we change the location of the third book? Say the locations of the 1-lb books are $x_1 = 0$, $x_2 = 2$, $x_3 = 3$, and $x_4 = 6$.
- Next, suppose that we place four books on the shelf, but of varying weights: at $x_1 = 0$ a 2-lb book, at $x_2 = 2$ a 3-lb book, at $x_3 = 4$ a 1-lb book, and at $x_4 = 6$ a 1-lb book. Use a weighted average of the locations to find \bar{x} , the balancing point of the shelf. How does the balancing point in this scenario compare to that found in (b)?
- What happens if we change the location of one of the books? Say that we keep everything the same in (d), except that $x_3 = 5$. How does \bar{x} change?

- f. What happens if we change the weight of one of the books? Say that we keep everything the same in (d), except that the book at $x_3 = 4$ now weighs 2 lbs. How does \bar{x} change?
- g. Experiment with a couple of different scenarios of your choosing where you move one of the books to the left, or you decrease the weight of one of the books.
- h. Write a couple of sentences to explain how adjusting the location of one of the books or the weight of one of the books affects the location of the balancing point of the shelf. Think carefully here about how your changes should be considered relative to the location of the balancing point \bar{x} of the current scenario.

Hint.

- a. Find the average location.
- b. Note that you are averaging 4 locations.
- c. Compare (b).
- d. If the book at location x_1 weighs 2 pounds, it's like there are two books at x_1 , so $2x_1$ must play a role in the average.
- e. Use a weighted average.
- f. Compare (e).
- g. Experiment.
- h. Compare the effects of moving locations left and adding weight to the left of the balancing point.

Answer.

- a. $\bar{x} = \frac{x_1+x_2}{2} = 3$.
- b. $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4} = 3$.
- c. $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4} = 2.75$.
- d. $\bar{x} = \frac{2x_1+3x_2+1x_3+1x_4}{7} = \frac{16}{7}$.
- e. $\bar{x} = \frac{2x_1+3x_2+1x_3+1x_4}{7} = \frac{17}{7}$.
- f. $\bar{x} = \frac{2x_1+3x_2+1x_3+2x_4}{7} = \frac{22}{7}$.
- g. Answers will vary.
- h. If we have an existing arrangement and balancing point, moving one of the locations to the left will move the balancing point to the left; similarly, moving one of the locations to the right will move the balancing point to the right. If instead we add weight to an existing location, if that location is left of the balancing point, the balancing point will move left; the behavior is similar if on the right.

Solution.

- The shelf will balance at $\bar{x} = \frac{x_1+x_2}{2} = \frac{0+6}{2} = 3$.
- The shelf will balance at the average of the locations: $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4} = \frac{0+2+4+6}{4} = 3$.
- Changing the location of x_3 to the left will move the average to the left, so the new location for \bar{x} is $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4} = \frac{0+2+3+6}{4} = \frac{11}{4} = 2.75$.
- We can think of the weighted books almost as if there were that *number* of 1-lb books at each location. The balancing point is given by the weighted average

$$\bar{x} = \frac{2x_1 + 3x_2 + 1x_3 + 1x_4}{7} = \frac{2 \cdot 0 + 3 \cdot 2 + 1 \cdot 4 + 1 \cdot 6}{7} = \frac{16}{7}.$$

Note particularly that we divide by the *total weight* of the books, not the total number.

- Here we find

$$\bar{x} = \frac{2x_1 + 3x_2 + 1x_3 + 1x_4}{7} = \frac{2 \cdot 0 + 3 \cdot 2 + 1 \cdot 5 + 1 \cdot 6}{7} = \frac{17}{7},$$

which moves the balancing point slightly to the right.

- In this situation, we find the balancing point moves significantly to the right:

$$\bar{x} = \frac{2x_1 + 3x_2 + 1x_3 + 2x_4}{7} = \frac{2 \cdot 0 + 3 \cdot 2 + 1 \cdot 4 + 2 \cdot 6}{7} = \frac{22}{7}.$$

- Try, for instance, having $x_2 = 1$ and $x_3 = 2$ and experiment with different weights.
- If we have an existing arrangement and balancing point, moving one of the locations to the left will move the balancing point to the left; similarly, moving one of the locations to the right will move the balancing point to the right. If instead we add weight to an existing location, if that location is left of the balancing point, the balancing point will move left; the behavior is similar if on the right.

6.3.3 Center of Mass

In Example 6.3.6, we saw that the balancing point of a system of point-masses¹ (such as books on a shelf) is found by taking a weighted average of their respective locations. In the example, we were computing the *center of mass* of a system of masses distributed along an axis, which is the balancing point of the axis on which the masses rest.

Center of Mass (point-masses)

For a collection of n masses m_1, \dots, m_n that are distributed along a single axis at the locations x_1, \dots, x_n , the center of mass is given by

$$\bar{x} = \frac{x_1m_1 + x_2m_2 + \dots + x_nm_n}{m_1 + m_2 + \dots + m_n}$$

Now consider a thin bar over which density is distributed continuously. If the density is constant, it is obvious that the balancing point of the bar is its midpoint. But if density is not

¹In the example, we actually used *weight* rather than *mass*. Since weight is proportional to mass, the computations for the balancing point result in the same location regardless of whether we use weight or mass. The gravitational constant is present in both the numerator and denominator of the weighted average.

constant, we must compute a weighted average. Let's say that the function $\rho(x)$ tells us the density distribution along the bar, measured in g/cm. If we slice the bar into small sections, we can think of the bar as holding a collection of adjacent point-masses. The mass m_i of a slice of thickness Δx at location x_i , is $m_i \approx \rho(x_i)\Delta x$.

If we slice the bar into n pieces, we can approximate its center of mass by

$$\bar{x} \approx \frac{x_1 \cdot \rho(x_1)\Delta x + x_2 \cdot \rho(x_2)\Delta x + \cdots + x_n \cdot \rho(x_n)\Delta x}{\rho(x_1)\Delta x + \rho(x_2)\Delta x + \cdots + \rho(x_n)\Delta x}.$$

Rewriting the sums in sigma notation, we have

$$\bar{x} \approx \frac{\sum_{i=1}^n x_i \cdot \rho(x_i)\Delta x}{\sum_{i=1}^n \rho(x_i)\Delta x}. \quad (6.3.1)$$

The greater the number of slices, the more accurate our estimate of the balancing point will be. The sums in Equation (6.3.1) can be viewed as Riemann sums, so in the limit as $n \rightarrow \infty$, we find that the center of mass is given by the quotient of two integrals.

Center of Mass (continuous mass distribution)

For a thin rod of density $\rho(x)$ distributed along an axis from $x = a$ to $x = b$, the center of mass of the rod is given by

$$\bar{x} = \frac{\int_a^b x \rho(x) dx}{\int_a^b \rho(x) dx}.$$

Note that the denominator of \bar{x} is the mass of the bar, and that this quotient of integrals is simply the continuous version of the weighted average of locations, x , along the bar.

Example 6.3.7. Consider a thin bar of length 20 cm whose density is distributed according to the function $\rho(x) = 4 + 0.1x$, where $x = 0$ represents the left end of the bar. Assume that ρ is measured in g/cm and x is measured in cm.

- Find the total mass, M , of the bar.
- Without doing any calculations, do you expect the center of mass of the bar to be equal to 10, less than 10, or greater than 10? Why?
- Compute \bar{x} , the exact center of mass of the bar.
- What is the average density of the bar?
- Now consider a different density function, given by $p(x) = 4e^{0.020732x}$, also for a bar of length 20 cm whose left end is at $x = 0$. Plot both $\rho(x)$ and $p(x)$ on the same axes. Without doing any calculations, which bar do you expect to have the greater center of mass? Why?
- Compute the exact center of mass of the bar described in (e) whose density function is $p(x) = 4e^{0.020732x}$. Check the result against the prediction you made in (e).

Hint.

- Recall $M = \int_a^b \rho(x) dx$.
- Think about whether there's more mass to the left of 10 or to the right of 10.
- Use the formula just developed in the preceding subsection of the text.
- Recall that when density, ρ is constant, $M = \rho \cdot V$.
- Think about how the two density functions distribute mass. Note that they appear to distribute the same total mass, as the areas under the two curves appear to be equal.
- Use $p(x)$ in place of $\rho(x)$ in standard calculations.

Answer.

- $M = \int_0^{20} 4 + 0.1x dx = 100$ g.
- Greater than 10.
- $\bar{x} = \frac{\int_0^{20} x(4+0.1x) dx}{\int_0^{20} 4+0.1x dx} = \frac{32}{3}$.
- 5 g/cm.
- Slightly to the right of the center of mass for $\rho(x)$.
- $\bar{x} = \frac{\int_0^{20} x4e^{0.020732x} dx}{\int_0^{20} 4e^{0.020732x} dx} \approx 10.6891$,

Solution.

- $M = \int_0^{20} 4 + 0.1x dx = 100$ g.
- The center of mass will be greater than 10 since the mass is more concentrated as x increases.
- Using the formula we developed, $\bar{x} = \frac{\int_0^{20} x\rho(x) dx}{\int_0^{20} \rho(x) dx}$, so

$$\bar{x} = \frac{\int_0^{20} x(4+0.1x) dx}{\int_0^{20} 4+0.1x dx} = \frac{\frac{3200}{3}}{100} = \frac{32}{3}$$
,
which is indeed just greater than 10.
- Recall that average density is given by mass divided by length, so since the bar is 20 cm long and its mass is 100 g, the average density is $100/20 = 5$ g/cm. This can also be computed using the formula for average value of a function:

$$\rho_{\text{AVG}[0,20]} = \frac{1}{20-0} \int_0^{20} \rho(x) dx = \frac{100}{20}.$$

- e. Since $p(x) = 4e^{0.020732x}$ lies below $\rho(x) = 4 + 0.1x$ up until where they intersect at about 17.53, after which $p(x) > \rho(x)$, we see that $p(x)$ distributes mass more to the right. It also appears as if the two curves may bound the same total area, and thus the generate the same total mass. If that is the case, we expect \bar{x} for $p(x)$ to lie slightly to the right of the center of mass for $\rho(x)$.

- f. We see that

$$\bar{x} = \frac{\int_0^{20} x 4e^{0.020732x} dx}{\int_0^{20} 4e^{0.020732x} dx} \approx 10.6891,$$

which is indeed slightly right of \bar{x} for $\rho(x)$.

6.3.4 Summary

- For an object of constant density D , with volume V and mass m , we know that $m = D \cdot V$.
- If an object with constant cross-sectional area (such as a thin bar) has its density distributed along an axis according to the function $\rho(x)$, then we can find the mass of the object between $x = a$ and $x = b$ by

$$m = \int_a^b \rho(x) dx.$$

- For a system of point-masses distributed along an axis, say m_1, \dots, m_n at locations x_1, \dots, x_n , the center of mass, \bar{x} , is given by the weighted average

$$\bar{x} = \frac{\sum_{i=1}^n x_i m_i}{\sum_{i=1}^n m_i}.$$

If instead we have mass continuously distributed along an axis, such as by a density function $\rho(x)$ for a thin bar of constant cross-sectional area, the center of mass of the portion of the bar between $x = a$ and $x = b$ is given by

$$\bar{x} = \frac{\int_a^b x \rho(x) dx}{\int_a^b \rho(x) dx}.$$

In each situation, \bar{x} represents the balancing point of the system of masses or of the portion of the bar.

6.3.5 Exercises

1. (Center of mass for a linear density function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Center of mass for a nonlinear density function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

3. (Interpreting the density of cars on a road) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Center of mass in a point-mass system) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. Let a thin rod of length a have density distribution function $\rho(x) = 10e^{-0.1x}$, where x is measured in cm and ρ in grams per centimeter.
- If the mass of the rod is 30 g, what is the value of a ?
 - For the 30g rod, will the center of mass lie at its midpoint, to the left of the midpoint, or to the right of the midpoint? Why?
 - For the 30g rod, find the center of mass, and compare your prediction in (b).
 - At what value of x should the 30g rod be cut in order to form two pieces of equal mass?

Answer.

- $a = -10 \ln(0.7) \approx 3.567$ cm.
- Left of the midpoint.
- $\bar{x} \approx \frac{50.3338}{30} \approx 1.687$.
- $q = -10 \ln(0.85) \approx 1.625$ cm.

Solution.

- To find mass we slice the rod horizontally into uniform slices of length Δx and approximate the density on the i th slice with the constant density $\rho(x_i)$. Then the approximate mass of the i th slice is $\rho(x_i)\Delta x$. Adding the approximations and taking the limit as the number of slices goes to infinity gives us the definite integral

$$\int_0^a \rho(x) dx$$

to represent the mass of the rod. Using the substitution $u = -0.1x$ and $du = -0.1 dx$ we have

$$\begin{aligned} \int_0^a \rho(x) dx &= \int_0^a 10e^{-0.1x} dx \\ &= \int_0^{-0.1a} -100e^u du \\ &= -100e^u \Big|_0^{-0.1a} \\ &= -100(e^{-0.1a} - 1) \\ &= 100(1 - e^{-0.1a}). \end{aligned}$$

So if the mass is 30 g, then $100(1 - e^{-0.1a}) = 30$, so $1 - e^{-0.1a} = 0.3$, and $e^{-0.1a} = 0.7$. It follows that $-0.1a = \ln(0.7)$, and thus, $a = -10 \ln(0.7) \approx 3.567$ cm.

- Since the density is a decreasing function, we should expect that the center of mass of the rod will lie to the left of the midpoint.

- c. To find the center of mass, we slice the rod horizontally into uniform slices of length Δx and approximate the mass on the i th slice with the constant density $\rho(x_i)$ to be $\rho(x_i)\Delta x$ as in (a). The position of the i th slice is approximately $x_i/$. Treating the rod as a collection of n discrete masses (the slices), the center of mass \bar{x} is approximately

$$\bar{x} \approx \frac{\sum_{i=1}^n x_i \rho(x_i) \Delta x}{\sum_{i=1}^n \rho(x_i) \Delta x}.$$

Letting the number of slices go to infinity gives us the exact location of the center of mass of the rod,

$$\bar{x} = \frac{\int_0^a x \rho(x) dx}{\int_0^a \rho(x) dx}.$$

The mass of the rod is 30 g, so now we just need to evaluate $\int_0^a x \rho(x) dx = \int_0^a 10xe^{-0.1x} dx$. Integrating by parts with $u = x$, $du = dx$, $dv = e^{-0.1x} dx$, and $v = -10e^{-0.1x}$ gives us

$$\begin{aligned} \int_0^a 10xe^{-0.1x} dx &= 10 \left(-10xe^{-0.1x} \Big|_0^a + \int_0^a -10e^{-0.1x} dx \right) \\ &= 10(-10ae^{-0.1a} + 100(1 - e^{-0.1a})) \\ &\approx 10(-10(3.567)ae^{-0.1(3.567)} + 100(1 - e^{-0.1(3.567)})) \\ &\approx 50.338. \end{aligned}$$

So the center of mass is $\bar{x} \approx \frac{50.338}{30} \approx 1.687$, which is less than midway across the 3.567 cm rod as expected.

- d. Each piece would have mass 15 grams. Let q be a value of x between 0 and a so the mass of the rod on the interval $[0, q]$ is 15. Using the substitution $u = -0.1x$ and $du = -0.1 dx$ we have that the mass of the rod on the interval $[0, q]$ is

$$\begin{aligned} \int_0^q \rho(x) dx &= \int_0^q 10e^{-0.1x} dx \\ &= \int_0^{-0.1q} -100e^u du \\ &= -100e^u \Big|_0^{-0.1q} \\ &= 100(1 - e^{-0.1q}) \end{aligned}$$

We want this mass to be 15 grams, so we find q such that $100(1 - e^{-0.1q}) = 15$. Solving for q in the usual way, we find $q = -10 \ln(0.85) \approx 1.625$ cm. Note that this is not the same as the center of mass. The center of mass takes into account the torque (determined by distance from the center of mass), which is not incorporated into the calculation of the point to split the object into two pieces of equal mass.

6. Consider two thin bars of constant cross-sectional area, each of length 10 cm, with respective mass density functions $\rho(x) = \frac{1}{1+x^2}$ and $p(x) = e^{-0.1x}$.

- a. Find the mass of each bar.

- b. Find the center of mass of each bar.
- c. Now consider a new 10 cm bar whose mass density function is $f(x) = \rho(x) + p(x)$.
- Explain how you can easily find the mass of this new bar with little to no additional work.
 - Similarly, compute $\int_0^{10} x f(x) dx$ as simply as possible, in light of earlier computations.
 - True or false: the center of mass of this new bar is the average of the centers of mass of the two earlier bars. Write at least one sentence to say why your conclusion makes sense.

Answer.

a. $M_1 = \arctan(10) \approx 1.47113; M_2 = 10 - 10e^{-1} \approx 6.32121$.

b. $\bar{x}_1 \approx 1.56857; \bar{x}_2 \approx 4.18023$.

c. (i)

$$M = \int_0^{10} \rho(x) dx + \int_0^{10} p(x) dx \approx 1.47113 + 6.32121 = 7.79234.$$

(ii)

$$\int_0^{10} x(\rho(x) + p(x)) dx = 28.73167.$$

(iii) False.

Solution.

- a. To find the mass of the first bar, we use the fact that $M = \int_0^a \rho(x) dx$, with $a = 10$ and $\rho(x) = \frac{1}{1+x^2}$. Thus, the mass of the first bar is

$$M_1 = \int_0^{10} \frac{1}{1+x^2} dx = \arctan(x)|_0^{10},$$

so $M_1 = \arctan(10) \approx 1.47113$.

Similar computations for the second bar shows that its mass is

$$M_2 = \int_0^{10} e^{-0.1x} dx = -10e^{-0.1x}|_0^{10},$$

so $M_2 = 10 - 10e^{-1} \approx 6.32121$.

- b. To find the center of mass of each bar, we use the standard formula:

$$\bar{x} = \frac{\int_0^{10} x \rho(x) dx}{\int_0^{10} \rho(x) dx}.$$

Having already evaluated the denominator for each mass distribution, we simply need to compute the numerator. It is straightforward to show that

$$\int_0^{10} x \frac{1}{1+x^2} dx = \frac{1}{2} \ln(101) \approx 2.30756,$$

so the first bar's center of mass is

$$\bar{x}_1 = \frac{\arctan(10)}{\frac{1}{2} \ln(101)} \approx \frac{2.30756}{1.47113} \approx 1.56857.$$

For the second bar,

$$\int_0^{10} x e^{-0.1x} dx = 100 - 200e^{-1} \approx 26.42411,$$

so its center of mass is

$$\bar{x}_2 = \frac{10 - 10e^{-1}}{100 - 200e^{-1}} \approx \frac{26.42411}{6.32121} \approx 4.18023.$$

c. Consider a new 10 cm bar whose mass density function is $f(x) = \rho(x) + p(x)$.

(i) The mass of this new bar is $M = \int_0^{10} (\rho(x) + p(x)) dx$. Using the additive property of the definite integral,

$$M = \int_0^{10} \rho(x) dx + \int_0^{10} p(x) dx \approx 1.47113 + 6.32121 = 7.79234.$$

(ii) Applying the same idea to $\int_0^{10} x f(x) dx$, we have

$$\int_0^{10} x(\rho(x) + p(x)) dx = \int_0^{10} x\rho(x) dx + \int_0^{10} xp(x) dx \approx 2.30756 + 26.42411 = 28.73167.$$

(iii) False. The center of mass of the new bar is

$$\bar{x} = \frac{\int_0^{10} x(\rho(x) + p(x)) dx}{\int_0^{10} (\rho(x) + p(x)) dx} \approx \frac{2.30756 + 26.42411}{1.47113 + 6.32121} = 3.68717$$

while the average of the two bars' centers of mass is

$$\frac{\bar{x}_1 + \bar{x}_2}{2} \approx \frac{\frac{2.30756}{1.47113} + \frac{26.42411}{6.32121}}{2} = 2.8744.$$

Although the computations prove the conclusion, the result is also intuitive: the average treats the two bars as equal contributors and doesn't take into account that one bar has much more mass than the other. Indeed, a weighted average is needed to correctly compute the center of mass, which is what our earlier computations in (i) and (ii) accomplish.

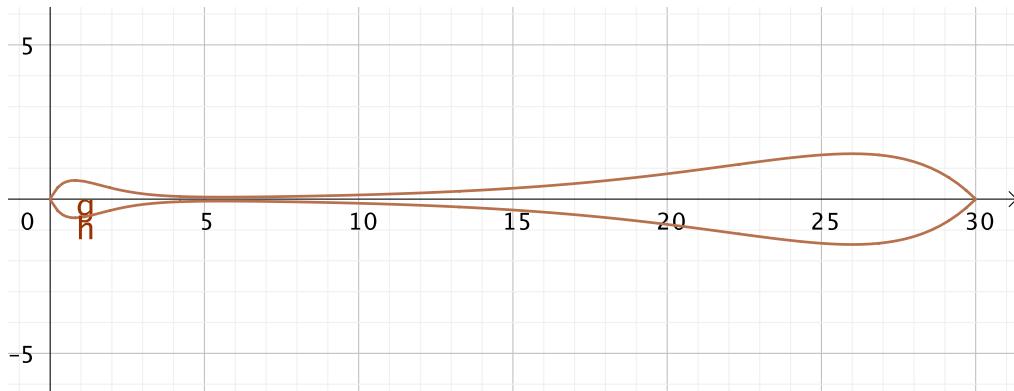
7. Consider the curve given by $y = f(x) = 2xe^{-1.25x} + (30 - x)e^{-0.25(30-x)}$.

- Plot this curve in the window $x = 0 \dots 30$, $y = 0 \dots 3$ (with constrained scaling so the units on the x and y axis are equal), and use it to generate a solid of revolution about the x -axis. Explain why this curve could generate a reasonable model of a baseball bat.
- Let x and y be measured in inches. Find the total volume of the baseball bat generated by revolving the given curve about the x -axis. Include units on your answer.

- c. Suppose that the baseball bat has constant weight density, and that the weight density is 0.6 ounces per cubic inch. Find the total weight of the bat whose volume you found in (b).
- d. Because the baseball bat does not have constant cross-sectional area, we see that the amount of weight concentrated at a location x along the bat is determined by the volume of a slice at location x . Explain why we can think about the function $\rho(x) = 0.6\pi f(x)^2$ (where f is the function given at the start of the problem) as being the weight density function for how the weight of the baseball bat is distributed from $x = 0$ to $x = 30$.
- e. Compute the center of mass of the baseball bat.

Answer.

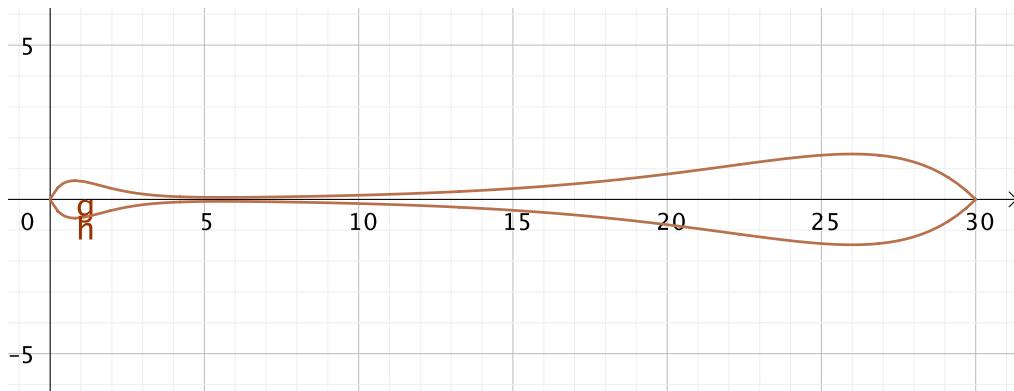
a.



- b. $V = \int_0^{30} \pi(2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 dx \approx 52.0666$ cubic inches.
- c. $W \approx 0.6 \cdot 52.0666 = 31.23996$ ounces.
- d. At a given x -location, the amount of weight concentrated there is approximately the weight density (0.6 ounces per cubic inch) times the volume of the slice, which is $V_{\text{slice}} \approx \pi f(x)^2$.
- e. $\bar{x} \approx 23.21415$

Solution.

a.



The plot shown above (with the curve reflected across the x -axis to generate the solid of revolution) shows why this solid of revolution is a reasonable approximation to a baseball bat (if perhaps a bit thin between $x = 3$ and $x = 9$).

- b. We use the disk method to find the volume of the solid of revolution. In the usual way, the volume of a typical slice is

$$V_{\text{slice}} \approx \pi(2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 \Delta x$$

and thus using a definite integral to sum the volumes of the slices, we get

$$V = \int_0^{30} \pi(2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 dx.$$

Using technology to evaluate the integral, we find that $V \approx 52.0666$ cubic inches.

- c. Given that the bat has (constant) weight density 0.6 ounces per cubic inch, it follows that the total weight of the bat is $W \approx 0.6 \cdot 52.0666 = 31.23996$ ounces.
- d. If we slice the baseball bat perpendicular to the horizontal axes, we get thin cylindrical disks as cross-sections. At a given x -location, the amount of weight concentrated there is approximately the weight density (0.6 ounces per cubic inch) times the volume of the slice, which is $V_{\text{slice}} \approx \pi f(x)^2$. Thus, we see that the amount of weight concentrated at a location x along the bat is given by $\rho(x) = 0.6\pi f(x)^2$, which we can therefore view as being the weight density function for how the weight of the baseball bat is distributed from $x = 0$ to $x = 30$.
- e. Even though we are dealing with “weight” instead of “mass”, we can compute the center of mass in the usual way since weight differs from mass only by the gravitational constant, and this constant is present in both the numerator and denominator of the center of mass. Thus, using the standard formula,

$$\bar{x} = \frac{\int_0^{30} x \rho(x) dx}{\int_0^{30} \rho(x) dx}$$

so

$$\begin{aligned} \bar{x} &= \frac{\int_0^{30} x \cdot 0.6\pi(2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 dx}{\int_0^{30} 0.6\pi(2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 dx} \\ &= \frac{\int_0^{30} x \cdot (2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 dx}{\int_0^{30} (2xe^{-1.25x} + (30-x)e^{-0.25(30-x)})^2 dx} \\ &\approx \frac{384.735}{16.5733} = 23.21415, \end{aligned}$$

so the center of mass is just over $2/3$ along the length of the bat.

6.4 Physics Applications: Work, Force, and Pressure

Motivating Questions

- How do we measure the work accomplished by a varying force that moves an object a certain distance?
- What is the total force exerted by water against a dam?
- How are both of the above concepts and their corresponding use of definite integrals similar to problems we have encountered in the past involving formulas such as “distance equals rate times time” and “mass equals density times volume”?

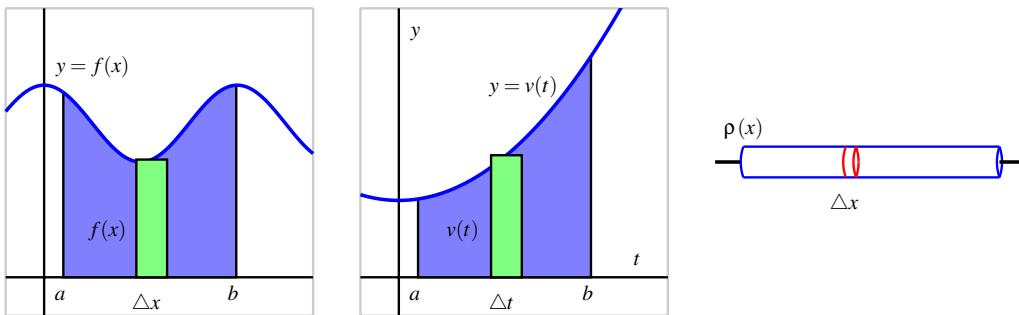


Figure 6.4.1: Three settings where we compute the accumulation of a varying quantity: the area under $y = f(x)$, the distance traveled by an object with velocity $y = v(t)$, and the mass of a bar with density function $y = \rho(x)$.

We have seen several different circumstances where the definite integral enables us to measure the accumulation of a quantity that varies, provided the quantity is approximately constant over small intervals. For instance, to find the area bounded by a nonnegative curve $y = f(x)$ and the x -axis on an interval $[a, b]$, we take a representative slice of width Δx that has area $A_{\text{slice}} = f(x)\Delta x$. As we let the width of the representative slice tend to zero, we find that the exact area of the region is

$$A = \int_a^b f(x) dx.$$

In a similar way, if we know the velocity $v(t)$ of a moving object and we wish to know the distance the object travels on an interval $[a, b]$ where $v(t)$ is nonnegative, we can use a definite integral to generalize the fact that $d = r \cdot t$ when the rate, r , is constant. On a short

time interval Δt , $v(t)$ is roughly constant, so for a small slice of time, $d_{\text{slice}} = v(t)\Delta t$. As the width of the time interval Δt tends to zero, the exact distance traveled is given by the definite integral

$$d = \int_a^b v(t) dt.$$

Finally, if we want to determine the mass of an object of non-constant density, because $M = D \cdot V$ (mass equals density times volume, provided that density is constant), we can consider a small slice of an object on which the density is approximately constant, and a definite integral may be used to determine the exact mass of the object. For instance, if we have a thin rod whose cross sections have constant density, but whose density is distributed along the x axis according to the function $y = \rho(x)$, it follows that for a small slice of the rod that is Δx thick, $M_{\text{slice}} = \rho(x)\Delta x$. In the limit as $\Delta x \rightarrow 0$, we then find that the total mass is given by

$$M = \int_a^b \rho(x) dx.$$

All three of these situations are similar in that we have a basic rule ($A = l \cdot w$, $d = r \cdot t$, $M = D \cdot V$) where one of the two quantities being multiplied is no longer constant; in each, we consider a small interval for the other variable in the formula, calculate the approximate value of the desired quantity (area, distance, or mass) over the small interval, and then use a definite integral to sum the results as the length of the small intervals is allowed to approach zero. It should be apparent that this approach will work effectively for other situations where we have a quantity that varies.

We next turn to the notion of *work*: from physics, a basic principle is that work is the product of force and distance. For example, if a person exerts a force of 20 pounds to lift a 20-pound weight 4 feet off the ground, the total work accomplished is

$$W = F \cdot d = 20 \cdot 4 = 80 \text{ foot-pounds.}$$

If force and distance are measured in English units (pounds and feet), then the units of work are *foot-pounds*. If we work in metric units, where forces are measured in Newtons and distances in meters, the units of work are *Newton-meters*.

Of course, the formula $W = F \cdot d$ only applies when the force is constant over the distance d . In Example 6.4.2, we explore one way that we can use a definite integral to compute the total work accomplished when the force exerted varies.

Example 6.4.2. A bucket is being lifted from the bottom of a 50-foot deep well; its weight (including the water), B , in pounds at a height h feet above the water is given by the function $B(h)$. When the bucket leaves the water, the bucket and water together weigh $B(0) = 20$ pounds, and when the bucket reaches the top of the well, $B(50) = 12$ pounds. Assume that the bucket loses water at a constant rate (as a function of height, h) throughout its journey from the bottom to the top of the well.

- Find a formula for $B(h)$.
- Compute the value of the product $B(5)\Delta h$, where $\Delta h = 2$ feet. Include units in your answer. Explain why this product represents the approximate work it took to move the bucket of water from $h = 5$ to $h = 7$.

- c. Is the value in (b) an over- or under-estimate of the actual amount of work it took to move the bucket from $h = 5$ to $h = 7$? Why?
- d. Compute the value of the product $B(22)\Delta h$, where $\Delta h = 0.25$ feet. Include units in your answer. What is the meaning of the value you found?
- e. More generally, what does the quantity $W_{\text{slice}} = B(h)\Delta h$ measure for a given value of h and a small positive value of Δh ?
- f. Evaluate the definite integral $\int_0^{50} B(h) dh$. What is the meaning of the value you find? Why?

6.4.1 Work

Because work is calculated by the rule $W = F \cdot d$ whenever the force F is constant, it follows that we can use a definite integral to compute the work accomplished by a varying force. For example, suppose that a bucket whose weight at height h is given by $B(h) = 12 + 8e^{-0.1h}$ is being lifted in a 50-foot well.

In contrast to the problem in Example 6.4.2, this bucket is not leaking at a constant rate; but because the weight of the bucket and water is not constant, we have to use a definite integral to determine the total work done in lifting the bucket. At a height h above the water, the approximate work to move the bucket a small distance Δh is

$$W_{\text{slice}} = B(h)\Delta h = (12 + 8e^{-0.1h})\Delta h.$$

Hence, if we let Δh tend to 0 and take the sum of all of the slices of work accomplished on these small intervals, it follows that the total work is given by

$$W = \int_0^{50} B(h) dh = \int_0^{50} (12 + 8e^{-0.1h}) dh.$$

While it is a straightforward exercise to evaluate this integral exactly using the First Fundamental Theorem of Calculus, in applied settings such as this one we will typically use computing technology. Here, it turns out that $W = \int_0^{50} (12 + 8e^{-0.1h}) dh \approx 679.461$ foot-pounds.

Our work in Example 6.4.2 and in the most recent discussion above employs the following important general principle.

For an object being moved in the positive direction along an axis with location x by a force $F(x)$, the total work to move the object from a to b is given by

$$W = \int_a^b F(x) dx.$$

Example 6.4.3. Consider the following situations in which a varying force accomplishes work.

- Suppose that a heavy rope hangs over the side of a cliff. The rope is 200 feet long and weighs 0.3 pounds per foot; initially the rope is fully extended. How much work is required to haul in the entire length of the rope? (Hint: set up a function $F(h)$ whose value is the weight of the rope remaining over the cliff after h feet have been hauled in.)
- A leaky bucket is being hauled up from a 100 foot deep well. When lifted from the water, the bucket and water together weigh 40 pounds. As the bucket is being hauled upward at a constant rate, the bucket leaks water at a constant rate so that it is losing weight at a rate of 0.1 pounds per foot. What function $B(h)$ tells the weight of the bucket after the bucket has been lifted h feet? What is the total amount of work accomplished in lifting the bucket to the top of the well?
- Now suppose that the bucket in (b) does not leak at a constant rate, but rather that its weight at a height h feet above the water is given by $B(h) = 25 + 15e^{-0.05h}$. What is the total work required to lift the bucket 100 feet? What is the average force exerted on the bucket on the interval $h = 0$ to $h = 100$?
- From physics, *Hooke's Law* for springs states that the amount of force required to hold a spring that is compressed (or extended) to a particular length is proportionate to the distance the spring is compressed (or extended) from its natural length. That is, the force to compress (or extend) a spring x units from its natural length is $F(x) = kx$ for some constant k (which is called the *spring constant*.) For springs, we choose to measure the force in pounds and the distance the spring is compressed in feet. Suppose that a force of 5 pounds extends a particular spring 4 inches ($1/3$ foot) beyond its natural length.
 - Use the given fact that $F(1/3) = 5$ to find the spring constant k .
 - Find the work done to extend the spring from its natural length to 1 foot beyond its natural length.
 - Find the work required to extend the spring from 1 foot beyond its natural length to 1.5 feet beyond its natural length.

Hint.

- Find a linear function $B(h)$ that satisfies $B(0) = 200 \cdot 0.3$ and $B(1) = 199 \cdot 0.3$
- The weight of the bucket is given by a linear function.
- Recall that work is given by a certain definite integral.
- Follow the rule for Hooke's Law and then use $W = \int_a^b F(x) dx$.

Answer.

- $W = \int_0^{200} 0.3(200 - h) dh = 6000$ foot-pounds.
- $W = \int_0^{100} (40 - 0.1h) dh = 3500$ foot-pounds.

- c. $B_{\text{AVG}}[0,100] \approx 25.9798$ pounds.
- d. For the given spring,
 - i. $k = 15$.
 - ii. $W = \int_0^1 15x \, dx = \frac{15}{2}$ foot-pounds.
 - iii. $W = \int_1^{1.5} 15x \, dx = 9.375$ foot-pounds.

Solution.

- a. When the full rope is extended over the cliff, the weight of the hanging rope is 200 pounds. Every foot of rope weighs 0.3 pounds; in addition, the quantity $(200 - h)$ measures the number of feet of rope hanging over the cliff when h feet of rope have been pulled in. This means that the function $F(h)$ whose value is the weight of the rope hanging over the cliff after h feet have been pulled in is given by $F(h) = 0.3(200 - h)$. When a small amount of rope, Δh , is pulled in, the work to move that slice of rope is given by

$$W_{\text{slice}} = F(h)\Delta h = 0.3(200 - h)\Delta h.$$

Thus, the total work to pull in the 200 feet of rope is given by

$$W = \int_0^{200} 0.3(200 - h) \, dh = 6000 \text{ foot-pounds.}$$

- b. Since $B(h)$ changes at a constant rate, B is a linear function. We know that $B(0) = 40$ and that B loses weight at a rate of 0.1 pounds per foot of rope hauled in. Thus, $B(h) = 40 - 0.1h$. It follows that the total work to lift the leaky bucket 100 feet is

$$W = \int_0^{100} (40 - 0.1h) \, dh = 3500 \text{ foot-pounds.}$$

- c. Given that $B(h) = 25 + 15e^{-0.05h}$ is the weight of the bucket when h feet of rope have been pulled in, it follows that the total work to move the bucket 100 feet is

$$W = \int_0^{100} (25 + 15e^{-0.05h}) \, dh \approx 2597.98 \text{ foot-pounds.}$$

In addition, the average force that was exerted on the bucket is precisely the average weight of the bucket, which is

$$B_{\text{AVG}}[0,100] = \frac{1}{100 - 0} \int_0^{100} (25 + 15e^{-0.05h}) \, dh \approx 25.9798 \text{ pounds.}$$

- d. Now considering the given spring,
 - i. Since $F(1/3) = 5$ and $F(x) = kx$ by Hooke's Law, it follows that $5 = k \cdot \frac{1}{3}$, and thus $k = 15$.
 - ii. Because the work to stretch the spring 1 foot is given by $W = \int_0^1 F(x) \, dx$, it follows $W = \int_0^1 15x \, dx = \frac{15}{2}$ foot-pounds.
 - iii. Here, $W = \int_1^{1.5} F(x) \, dx = \int_1^{1.5} 15x \, dx = 9.375$ foot-pounds.

6.4.2 Work: Pumping Liquid from a Tank

In certain geographic locations where the water table is high, residential homes with basements have a peculiar feature: in the basement, one finds a large hole in the floor, and in the hole, there is water. For example, in Figure 6.4.4 we see a *sump crock*¹. A sump crock provides an outlet for water that may build up beneath the basement floor to prevent flooding the basement.

In the crock we see a floating pump. This pump is activated by elevation, so when the water level reaches a particular height, the pump turns on and pumps water out of the crock, hence relieving the water buildup beneath the foundation. One of the questions we'd like to answer is: how much work does a sump pump accomplish?



Figure 6.4.4: A sump crock.

Example 6.4.5. Suppose that a sump crock has the shape of a frustum of a cone, as pictured in Figure 6.4.6. The crock has a diameter of 3 feet at its surface, a diameter of 1.5 feet at its base, and a depth of 4 feet. In addition, suppose that the sump pump is set up so that it pumps the water vertically up a pipe to a drain that is located at ground level just outside a basement window. To accomplish this, the pump must send the water to a location 9 feet above the surface of the sump crock. How much work is required to empty the sump crock if it is initially completely full?

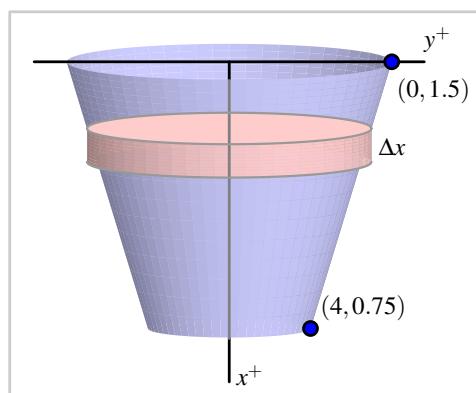


Figure 6.4.6: A sump crock with approximately cylindrical cross-sections.

¹Image credit to www.warreninspect.com/basement-moisture.

Solution. It is helpful to think of the depth below the surface of the crock as being the independent variable, so we let the positive x -axis point down, and the positive y -axis to the right, as pictured in the figure. Because the pump sits on the surface of the water, it makes sense to think about the pump moving the water one “slice” at a time, where it takes a thin slice from the surface, pumps it out of the tank, and then proceeds to pump the next slice below.

Each slice of water is cylindrical in shape. We see that the radius of each slice varies according to the linear function $y = f(x)$ that passes through the points $(0, 1.5)$ and $(4, 0.75)$, where x is the depth of the particular slice in the tank; it is a straightforward exercise to find that $f(x) = 1.5 - 0.1875x$. Now we think about the problem in several steps:

- a determining the volume of a typical slice;
- b finding the weight² of a typical slice (and thus the force that must be exerted on it);
- c deciding the distance that a typical slice moves;
- d and computing the work to move a representative slice.

Once we know the work it takes to move one slice, we use a definite integral over an appropriate interval to find the total work.

Consider a representative cylindrical slice at a depth of x feet below the top of the crock. The approximate volume of that slice is given by

$$V_{\text{slice}} = \pi f(x)^2 \Delta x = \pi(1.5 - 0.1875x)^2 \Delta x.$$

Since water weighs 62.4 lb/ft³, the approximate weight of a representative slice is

$$F_{\text{slice}} = 62.4 \cdot V_{\text{slice}} = 62.4\pi(1.5 - 0.1875x)^2 \Delta x.$$

This is also the approximate force the pump must exert to move the slice.

Because the slice is located at a depth of x feet below the top of the crock, the slice being moved by the pump must move x feet to get to the level of the basement floor, and then, as stated in the problem description, another 9 feet to reach the drain at ground level. Hence, the total distance a representative slice travels is

$$d_{\text{slice}} = x + 9.$$

Finally, the work to move a representative slice is given by

$$W_{\text{slice}} = F_{\text{slice}} \cdot d_{\text{slice}} = 62.4\pi(1.5 - 0.1875x)^2 \Delta x \cdot (x + 9).$$

We sum the work required to move slices throughout the tank (from $x = 0$ to $x = 4$), let $\Delta x \rightarrow 0$, and hence

$$W = \int_0^4 62.4\pi(1.5 - 0.1875x)^2(x + 9) dx.$$

When evaluated using appropriate technology, the integral shows that the total work is $W = 3463.2\pi$ foot-pounds.

²We assume that the weight density of water is 62.4 pounds per cubic foot.

The preceding example demonstrates the standard approach to finding the work required to empty a tank filled with liquid. The main task in each such problem is to determine the volume of a representative slice, followed by the force exerted on the slice, as well as the distance such a slice moves. In the case where the units are metric, there is one key difference: in the metric setting, rather than weight, we normally first find the mass of a slice. For instance, if distance is measured in meters, the mass density of water is 1000 kg/m^3 . In that setting, we can find the mass of a typical slice (in kg). To determine the force required to move it, we use $F = ma$, where m is the object's mass and a is the gravitational constant 9.81 N/kg^3 . That is, in metric units, the weight density of water is 9810 N/m^3 .

Example 6.4.7. In each of the following problems, determine the total work required to accomplish the described task. In parts (b) and (c), a key step is to find a formula for a function that describes the curve that forms the side boundary of the tank.

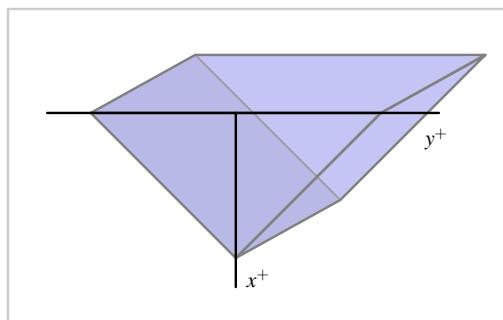


Figure 6.4.8: A trough with triangular ends, as described in Example 6.4.7, part (c).

- Consider a vertical cylindrical tank of radius 2 meters and depth 6 meters. Suppose the tank is filled with 4 meters of water of mass density 1000 kg/m^3 , and the top 1 meter of water is pumped over the top of the tank.
- Consider a hemispherical tank with a radius of 10 feet. Suppose that the tank is full to a depth of 7 feet with water of weight density 62.4 pounds/ft^3 , and the top 5 feet of water are pumped out of the tank to a tanker truck whose height is 5 feet above the top of the tank.
- Consider a trough with triangular ends, as pictured in Figure 6.4.8, where the tank is 10 feet long, the top is 5 feet wide, and the tank is 4 feet deep. Say that the trough is full to within 1 foot of the top with water of weight density 62.4 pounds/ft^3 , and a pump is used to empty the tank until the water remaining in the tank is 1 foot deep.

Hint.

- Note that slices of water at constant depth are all cylinders and all have the same radius. Also, remember to convert mass to weight when computing force.

- b. Recall that the top half of a circle of radius 10 centered at the origin has equation $y = \sqrt{100 - x^2}$.
- c. The equation of the line that determines the right side of the front face of the tank is $y = \frac{5}{2} - \frac{5}{8}x$. Note that a slice of water at constant depth is a rectangular slab whose width is $2y$.

Answer.

a.

$$W = \int_0^1 9.8 \cdot 4000\pi \cdot x \, dx = 19600\pi \text{ newton-meters.}$$

b.

$$W = \int_3^8 62.4\pi(100 - x^2)(x + 5) \, dx \approx 673593 \text{ foot-pounds.}$$

c.

$$W = \int_1^3 62.4(50 - \frac{25}{2}x)x \, dx = 5720 \text{ foot-pounds.}$$

Solution.

- a. A typical slice is a thin cylinder of radius 2 and thickness Δx . The volume of such a cylinder is $V_{\text{slice}} = \pi(2)^2\Delta x$, and its weight is thus $F_{\text{slice}} = 1000 \cdot 9.8 \cdot V_{\text{slice}} = 4000\pi\Delta x$. Finally, the distance such a slice travels to the top of the tank is $d_{\text{slice}} = x$. If we only pump out the top 1 m of water, it follows that the total work is

$$W = \int_0^1 9.8 \cdot 4000\pi \cdot x \, dx = 19600\pi \text{ newton-meters.}$$

- b. Viewing the tank as a solid of revolution, the tank is generated by the function $R(x) = \sqrt{100 - x^2}$, and a typical slice has volume $V_{\text{slice}} = \pi(\sqrt{100 - x^2})^2\Delta x$. The force to move such a slice is its weight, $F_{\text{slice}} = 62.4 \cdot V_{\text{slice}} = 62.4\pi(100 - x^2)\Delta x$. Each slice has to move x feet to clear the top of the tank, and then an additional 5 feet up to the truck, so the distance each slice moves is $d_{\text{slice}} = x + 5$. Finally, since the tank is full to a depth of 7 feet, and we wish to pump out the top 5 feet, x ranges from $x = 3$ (the top of the water) to $x = 8$ (the point at which there are 2 feet remaining in the tank). Hence, the total work done is

$$W = \int_3^8 62.4\pi(100 - x^2)(x + 5) \, dx \approx 673593 \text{ foot-pounds.}$$

- c. The line that bounds the right edge of the triangular face of the tank lying in the first quadrant is $y = \frac{5}{2} - \frac{5}{8}x$. If we take a slice of water that lies at depth x , note that the slice is rectangular with thickness Δx , length 10 (the tank is 10 feet long), and width $2(\frac{5}{2} - \frac{5}{8}x)$. Thus, the volume of a typical slice is

$$V_{\text{slice}} = 10 \cdot 2(\frac{5}{2} - \frac{5}{8}x) \cdot \Delta x = (50 - \frac{25}{4}x)\Delta x,$$

and the force that has to be exerted to move such a slice is

$$F_{\text{slice}} = 62.4 \cdot V_{\text{slice}} = 62.4\left(50 - \frac{25}{2}x\right)\Delta x.$$

The distance a typical slice moves when pumped to the top of the tank is $d_{\text{slice}} = x$, and since the tank is full to a depth of 3 feet (within 1 foot of the top) and we are going to empty the tank until 1 foot of water remains in it, we see that we need to integrate from $x = 1$ to $x = 3$, so

$$W = \int_1^3 62.4\left(50 - \frac{25}{2}x\right)x \, dx = 5720 \text{ foot-pounds.}$$

6.4.3 Force due to Hydrostatic Pressure

When building a dam, engineers need to know how much force water will exert against the face of the dam. This force comes from water pressure. The pressure a force exerts on a region is measured in units of force per unit of area: for example, the air pressure in a tire is often measured in pounds per square inch (PSI). Hence, we see that the general relationship is given by

$$P = \frac{F}{A}, \text{ or } F = P \cdot A,$$

where P represents pressure, F represents force, and A the area of the region being considered. Of course, in the equation $F = PA$, we assume that the pressure is constant over the entire region A .

We know from experience that the deeper one dives underwater while swimming, the greater the pressure exerted by the water. This is because at a greater depth, there is more water right on top of the swimmer: it is the force that “column” of water exerts that determines the pressure the swimmer experiences. The total water pressure is found by computing the total weight of the column of water that lies above a region of area 1 square foot at a fixed depth. At a depth of d feet, a rectangular column has volume $V = 1 \cdot 1 \cdot d \text{ ft}^3$, so the corresponding weight of the water overhead is $62.4d$. This is the amount of force being exerted on a 1 square foot region at a depth d feet underwater, so the pressure exerted by water at depth d is $P = 62.4d$ (lbs/ft²).

Because pressure is force per unit area, or $P = \frac{F}{A}$, we can compute the total force from a variable pressure by integrating $F = PA$.

Example 6.4.9. Consider a trapezoid-shaped dam that is 60 feet wide at its base and 90 feet wide at its top, and assume the dam is 25 feet tall with water that rises to within 5 feet of the top of its face. Water weighs 62.4 pounds per cubic foot. How much force does the water exert against the dam?

Solution. First, we sketch a picture of the dam, as shown in Figure 6.4.10. Note that, as in problems involving the work to pump out a tank, we let the positive x -axis point down.

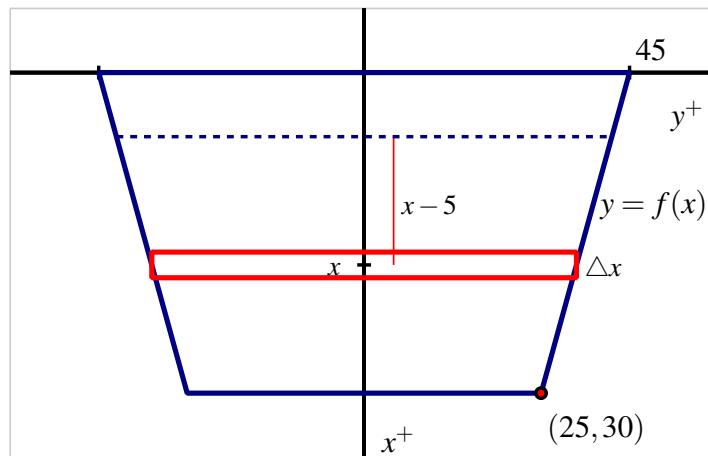


Figure 6.4.10: A trapezoidal dam that is 25 feet tall, 60 feet wide at its base, 90 feet wide at its top, with the water line 5 feet down from the top of its face.

Pressure is constant at a fixed depth, so we consider a slice of water at constant depth on the face, as shown in the figure. The area of this slice is approximately the area of the rectangle pictured. Since the width of that rectangle depends on the variable x , we find a formula for the line that represents one side of the dam. It is straightforward to find that $y = 45 - \frac{3}{5}x$. Hence, the approximate area of a representative slice is

$$A_{\text{slice}} = 2f(x)\Delta x = 2\left(45 - \frac{3}{5}x\right)\Delta x.$$

At any point on this slice, the depth is approximately constant, so the pressure can be considered constant. Because the water rises to within 5 feet of the top of the dam, the depth of any point on the representative slice is approximately $(x - 5)$. Now, since pressure is given by $P = 62.4d$, we have that at any point on the slice

$$P_{\text{slice}} = 62.4(x - 5).$$

Knowing both the pressure and area, we can find the force the water exerts on the slice. Using $F = PA$, it follows that

$$F_{\text{slice}} = P_{\text{slice}} \cdot A_{\text{slice}} = 62.4(x - 5) \cdot 2\left(45 - \frac{3}{5}x\right)\Delta x.$$

Finally, we use a definite integral to sum the forces over the appropriate range of x -values. Since the water rises to within 5 feet of the top of the dam, we start at $x = 5$ and take slices all the way to the bottom of the dam, where $x = 30$. Hence,

$$F = \int_{x=5}^{x=25} 62.4(x - 5) \cdot 2\left(45 - \frac{3}{5}x\right) dx.$$

Using technology to evaluate the integral, we find $F = 848640$ pounds.

Example 6.4.11. In each of the following problems, determine the total force exerted by water against the surface that is described.

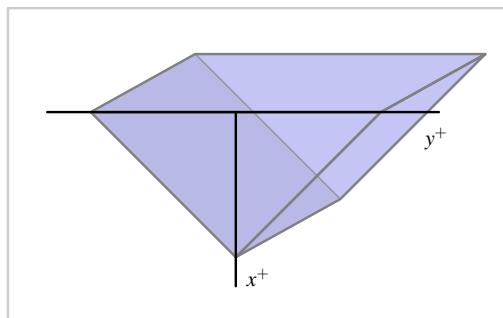


Figure 6.4.12: A trough with triangular ends, as described in Example 6.4.11, part (c).

- Consider a rectangular dam that is 100 feet wide and 50 feet tall, and suppose that water presses against the dam all the way to the top.
- Consider a semicircular dam with a radius of 30 feet. Suppose that the water rises to within 10 feet of the top of the dam.
- Consider a trough with triangular ends, as pictured in Figure 6.4.12, where the tank is 10 feet long, the top is 5 feet wide, and the tank is 4 feet deep. Say that the trough is full to within 1 foot of the top with water of weight density 62.4 pounds/ft³. How much force does the water exert against one of the triangular ends?

Hint.

- Put the center of the top of the dam at the origin and let the positive x -axis point down.
- With the positive x -axis pointing down, the equation for the edge of the semi-circular dam is $y = \sqrt{900 - x^2}$.
- The equation for the line that bounds the right side of the trough is $y = -\frac{5}{8}x + 2.5$.

Answer.

- $F = \int_{x=0}^{x=50} (6240x)dx = 7\,800\,000 \text{ pounds}.$
- $F = \int_{x=4}^{x=30} 124.8(x-4)\sqrt{900-x^2}dx = 800\,244 \text{ pounds}.$
- $F = \int_{x=1}^{x=4} 62.4(x-1)(5-1.25x)dx = 351 \text{ pounds}.$

Solution.

- a. We let the positive x -axis point down and let x be the distance from the top of the dam. The area of a typical horizontal slice (in square feet) is

$$A_{\text{slice}} = 100\Delta x.$$

The pressure (in pounds per square foot) at depth x is given by $P_{\text{slice}} = 62.5x$. So the force on a typical slice (in pounds) is

$$F_{\text{slice}} = (62.5x)(100\Delta x) = 6250x\Delta x.$$

So the force of the water against the dam is

$$F = \int_{x=0}^{x=50} (6250x)dx = 7800000 \text{ pounds}.$$

- b. We let the positive x -axis point down and let x be the distance from the top of the dam. We place the top of the dam along the x -axis with the center at the origin. So the equation for the semicircle is $y = \sqrt{30^2 - x^2} = \sqrt{900 - x^2}$. The area of a typical horizontal slice (in square feet) is

$$A_{\text{slice}} = 2y\Delta x = 2\sqrt{900 - x^2}\Delta x.$$

At a given value of x , the depth of the water is $(x - 4)$. So the pressure (in pounds per square foot) at x is given by $P_{\text{slice}} = 62.4(x - 4)$. So the force on a typical slice (in pounds) is

$$F_{\text{slice}} = 62.4(x - 4)(2\sqrt{900 - x^2}\Delta x) = 124.8(x - 4)\sqrt{900 - x^2}\Delta x.$$

So the force of the water against the dam is

$$F = \int_{x=4}^{x=30} 124.8(x - 4)\sqrt{900 - x^2}dx = 800244 \text{ pounds}.$$

- c. We let the positive x -axis point down and let x be the distance from the top of the dam. We place the top of the trough along the x -axis with the center at the origin. We need the equation for the straight line from the point $(0, 2.5)$ to the point $(4, 0)$. The equation for this line is $y = -\frac{5}{8}x + 2.5 = 2.5 - 0.625x$. The area of a typical horizontal slice (in square feet) is

$$A_{\text{slice}} = 2y\Delta x = 2(2.5 - 0.625x)\Delta x = (5 - 1.25x)\Delta x.$$

At a given value of x , the depth of the water is $(x - 1)$. So the pressure (in pounds per square foot) at x is given by $P_{\text{slice}} = 62.4(x - 1)$. So the force on a typical slice (in pounds) is

$$F_{\text{slice}} = 62.4(x - 1)(5 - 1.25x)\Delta x = 62.4(x - 1)(5 - 1.25x)\Delta x.$$

So the force of the water against the trough is

$$F = \int_{x=1}^{x=4} 62.4(x - 1)(5 - 1.25x)dx = 351 \text{ pounds}.$$

Although there are many different formulas involving work, force, and pressure, the fundamental ideas behind these problems are similar to others we've encountered in applications of the definite integral. We slice the quantity of interest into more manageable pieces and then use a definite integral to add them up.

6.4.4 Summary

- To measure the work done by a varying force in moving an object, we divide the problem into pieces on which we can use the formula $W = F \cdot d$, and then use a definite integral to sum the work done on each piece.
- To find the total force exerted by water against a dam, we use the formula $F = P \cdot A$ to measure the force exerted on a slice that lies at a fixed depth, and then use a definite integral to sum the forces across the appropriate range of depths.
- Because work is computed as the product of force and distance (provided force is constant), and the force water exerts on a dam can be computed as the product of pressure and area (provided pressure is constant), problems involving these concepts are similar to earlier problems we did using definite integrals to find distance (via “distance equals rate times time”) and mass (“mass equals density times volume”).

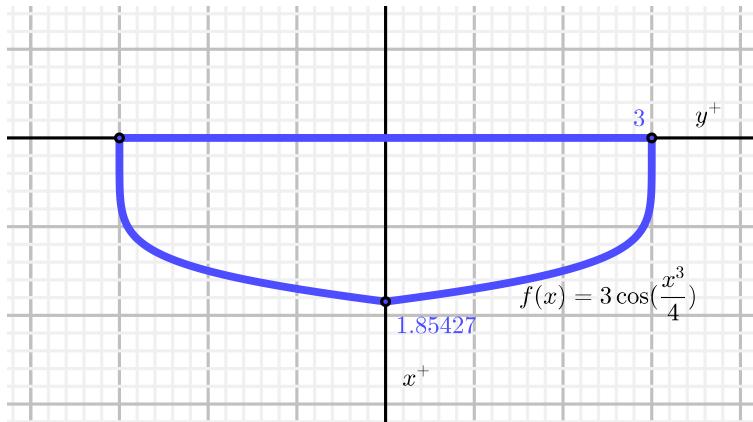
6.4.5 Exercises

1. (Work to empty a conical tank) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Work to empty a cylindrical tank) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Work to empty a rectangular pool) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Work to empty a cylindrical tank to differing heights) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Force due to hydrostatic pressure) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. Consider the curve $f(x) = 3 \cos(\frac{x^3}{4})$ and the portion of its graph that lies in the first quadrant between the y -axis and the first positive value of x for which $f(x) = 0$. Let R denote the region bounded by this portion of f , the x -axis, and the y -axis. Assume that x and y are each measured in feet.
 - a. Picture the coordinate axes rotated 90 degrees clockwise so that the positive x -axis points straight down, and the positive y -axis points to the right. Suppose that R is rotated about the x axis to form a solid of revolution, and we consider this solid as a storage tank. Suppose that the resulting tank is filled to a depth of 1.5 feet with water weighing 62.4 pounds per cubic foot. Find the amount of work required to lower the water in the tank until it is 0.5 feet deep, by pumping the water to the top of the tank.
 - b. Again picture the coordinate axes rotated 90 degrees clockwise so that the positive x -axis points straight down, and the positive y -axis points to the right. Suppose that R , together with its reflection across the x -axis, forms one end of a storage tank that is 10 feet long. Suppose that the resulting tank is filled completely with water weighing 62.4 pounds per cubic foot. Find a formula for a function that tells the amount of work required to lower the water by h feet.

- c. Suppose that the tank described in (b) is completely filled with water. Find the total force due to hydrostatic pressure exerted by the water on one end of the tank.

Answer.

a.



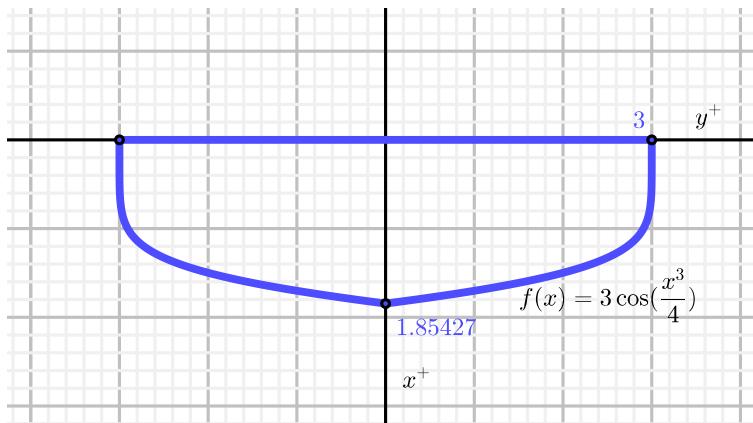
$$W \approx 1646.79 \text{ foot-pounds.}$$

$$b. W = \int_0^h 3744x \cos\left(\frac{x^3}{4}\right) dx.$$

$$c. F \approx 462.637 \text{ pounds.}$$

Solution.

- a. First, we construct a picture of the tank as described. Below we see a sketch of $f(x) = 3 \cos\left(\frac{x^3}{4}\right)$ in the setting where the positive x -axis points down; note that this function first crosses the x -axis at about 1.84527. We reflect the curve from the first quadrant across the positive x -axis to generate a solid of revolution. Cross-sections perpendicular to the x -axis have thickness Δx and are circular disks.



We now suppose that the resulting tank is filled to a depth of 1.5 feet with water weighing 62.4 pounds per cubic foot. Since the bottom of the tank is at 1.84527, being full to a depth of 1.5 feet means that there is water from $x = 0.34527$ to $x = 1.84527$. To find the work done to pump the water until 0.5 feet remain in the tank (i.e. to a level where

$x = 1.34527$), we have to first find the weight of a typical slice, which requires us to know the volume of such a slice. Observe that for a slice at depth x ,

$$V_{\text{slice}} \approx \pi \left(3 \cos \left(\frac{x^3}{4} \right) \right)^2 \Delta x.$$

Since water weighs 62.4 pounds per cubic foot, the weight (and thus force to be exerted) of such a slice is

$$F_{\text{slice}} \approx 62.4 \cdot 9 \cdot \pi \cos^2 \left(\frac{x^3}{4} \right) \Delta x.$$

Finally, the work done to pump such a slice to the top of the tank requires the slice to move x units, and thus work for that slice is

$$W_{\text{slice}} \approx x \cdot 561.6\pi \cos^2 \left(\frac{x^3}{4} \right) \Delta x.$$

To find the total work accomplished, we use a definite integral to sum the work done on each slice from $x = 0.34527$ (the top of the water when the tank is initially full to 1.84527 feet) to $x = 1.34527$ (the height of the water after 1 foot has been emptied). Thus,

$$W = \int_{0.34527}^{1.34527} 561.6\pi \cos^2 \left(\frac{x^3}{4} \right) dx.$$

Evaluating the integral, we find $W \approx 1646.79$ foot-pounds.

- b. If instead we envision the tank not as a solid of revolution but as a trough, where the two-dimensional figure above forms one end of a tank that is 10 feet long, we now have a similar problem but with slices that are thin rectangular slabs, rather than thin cylindrical disks. Note that each such slice is $2f(x)$ wide, 10 feet long, and Δx thick. Thus, for a slice at depth x ,

$$V_{\text{slice}} \approx 2 \left(3 \cos \left(\frac{x^3}{4} \right) \right) \cdot 10 \cdot \Delta x.$$

Since water weighs 62.4 pounds per cubic foot, the weight (and thus force to be exerted) of such a slice is

$$F_{\text{slice}} \approx 62.4 \cdot 60 \cos \left(\frac{x^3}{4} \right) \Delta x.$$

Finally, the work done to pump such a slice to the top of the tank requires the slice to move x units, and thus work for that slice is

$$W_{\text{slice}} \approx x \cdot 3744 \cos \left(\frac{x^3}{4} \right) \Delta x.$$

Assuming that the tank is completely full, to lower the tank by h feet, we consider slices from $x = 0$ to $x = h$, and thus using a definite integral to sum the work required for each slice, we find that the total work is

$$W = \int_0^h 3744x \cos \left(\frac{x^3}{4} \right) dx.$$

- c. For the trough described in (b), we now seek to find the total hydrostatic force on one of the ends of the tank when it is completely full. First, we note that a typical slice at location x will be at a depth of x feet below the water and have area

$$A_{\text{slice}} \approx 2f(x)\Delta x = 2 \left(3 \cos\left(\frac{x^3}{4}\right) \right) \Delta x.$$

At any point on this slice, which again lies at depth x , the pressure exerted is

$$P_{\text{slice}} \approx 62.4x$$

since pressure is the weight density of water times depth. Finally, since $F = P \cdot A$, the total force water is exerting on this slice is

$$F_{\text{slice}} = P_{\text{slice}} \cdot A_{\text{slice}} \approx 62.4x \cdot 6 \cos\left(\frac{x^3}{4}\right) \Delta x.$$

Integrating to sum the total force, we find that

$$F = \int_0^{1.84527} 374.4x \cos\left(\frac{x^3}{4}\right) dx \approx 462.637$$

pounds.

7. A cylindrical tank, buried on its side, has radius 3 feet and length 10 feet. It is filled completely with water whose weight density is 62.4 lbs/ft³, and the top of the tank is two feet underground.

- Set up, but do not evaluate, an integral expression that represents the amount of work required to empty the top half of the water in the tank to a truck whose tank lies 4.5 feet above ground.
- With the tank now only half-full, set up, but do not evaluate an integral expression that represents the total force due to hydrostatic pressure against one end of the tank.

Answer.

a. $W = 5904(19\pi - 8) \approx 305179.3$ foot-pounds.

b.

$$F \approx 1123.2$$

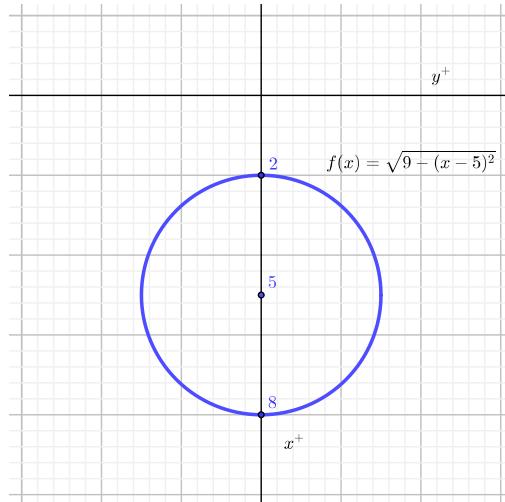
pounds.

Solution.

- First, we set up axes and a sketch for one of the ends of the cylindrical tank. It is simplest to let ground level correspond to the (horizontal) y -axis with the positive x -axis pointing straight down, so that $x = 0$ corresponds to ground level and the value of x tells us the depth of a particular location. The circular end of the tank has radius 3 and is centered at the point $(5, 0)$ since its top is 2 feet underground. Thus, the equation of that circle is

$$(x - 5)^2 + y^2 = 3^2.$$

Solving for y , we find the equation of one half of the circle, the half that is pictured at right in the figure below: $y = f(x) = \sqrt{9 - (x - 5)^2}$.



Next, for a slice at depth x , its shape is a rectangular slab with thickness Δx . Its length is 10, and its width is $2f(x) = 2\sqrt{9 - (x - 5)^2}$. Thus, the volume of such a slice is

$$V_{\text{slice}} \approx 10 \cdot 2\sqrt{9 - (x - 5)^2} \cdot \Delta x.$$

With the water weighing 62.4 pounds per cubic foot, the weight of (and thus force to be exerted upon) a typical slice is

$$F_{\text{slice}} \approx 262.4 \cdot 20\sqrt{9 - (x - 5)^2} \Delta x.$$

To move such a slice from its depth x feet below the ground to a tank on a truck that's 4.5 feet above ground, each slice must travel $x + 4.5$ feet, and thus the work to move the slice is

$$W_{\text{slice}} \approx (x + 4.5)5248\sqrt{9 - (x - 5)^2} \Delta x.$$

Finally, to empty half the tank, we need to consider slices from $x = 2$ to $x = 5$, and thus using a definite integral to sum the work required for each slice, we get

$$W = \int_2^5 5248(x + 4.5)\sqrt{9 - (x - 5)^2} dx = 5904(19\pi - 8) \approx 305179.3$$

foot-pounds.

- b. To find the hydrostatic force due to water pressure that is exerted on one end of the tank by the remaining water, we first find the area of a rectangular slice at a location x , which is

$$A_{\text{slice}} \approx 2f(x)\Delta x = 2\sqrt{9 - (x - 5)^2}\Delta x.$$

We note that such a slice lies between $x = 5$ and $x = 8$, which are the values of x that represent the depths of remaining water from top to bottom. At an x in this interval, the slice's depth below the surface of the water is $x - 5$. Thus, since pressure is weight-density times depth, it follows that

$$P_{\text{slice}} \approx 62.4(x - 5).$$

Finally, since $F = P \cdot A$, the total force water is exerting on this slice is

$$F_{\text{slice}} = P_{\text{slice}} \cdot A_{\text{slice}} \approx 62.4(x - 5) \cdot 2\sqrt{9 - (x - 5)^2} \Delta x.$$

Integrating to sum the total force, we find

$$F = \int_5^8 124.8(x - 5)\sqrt{9 - (x - 5)^2} dx \approx 1123.2$$

pounds.

6.5 Improper Integrals

Motivating Questions

- What are improper integrals and why are they important?
- What does it mean to say that an improper integral converges or diverges?
- What are some typical improper integrals that we can classify as convergent or divergent?

Another important application of the definite integral measures the likelihood of certain events. For instance, consider a company that manufactures incandescent light bulbs. Based on a large volume of test results, they have determined that the fraction of light bulbs that fail between times $t = a$ and $t = b$ of use (where t is measured in months) is given by

$$\int_a^b 0.3e^{-0.3t} dt.$$

For example, the fraction of light bulbs that fail during their third month of use is given by

$$\begin{aligned} \int_2^3 0.3e^{-0.3t} dt &= -e^{-0.3t} \Big|_2^3 \\ &= -e^{-0.9} + e^{-0.6} \\ &\approx 0.1422. \end{aligned}$$

Thus about 14.22% of all lightbulbs fail between $t = 2$ and $t = 3$. Clearly we could adjust the limits of integration to measure the fraction of light bulbs that fail during any time period of interest.

Example 6.5.1. A company with a large customer base has a call center that receives thousands of calls a day. After studying the data that represents how long callers wait for assistance, they find that the function $p(t) = 0.25e^{-0.25t}$ models the time customers wait in the following way: the fraction of customers who wait between $t = a$ and $t = b$ minutes is given by

$$\int_a^b p(t) dt.$$

Use this information to answer the following questions.

- Determine the fraction of callers who wait between 5 and 10 minutes.
- Determine the fraction of callers who wait between 10 and 20 minutes.
- Next, let's study the fraction who wait up to a certain number of minutes:

- i. What is the fraction of callers who wait between 0 and 5 minutes?
- ii. What is the fraction of callers who wait between 0 and 10 minutes?
- iii. Between 0 and 15 minutes? Between 0 and 20?
- d. Let $F(b)$ represent the fraction of callers who wait between 0 and b minutes. Find a formula for $F(b)$ that involves a definite integral, and then use the First FTC to find a formula for $F(b)$ that does not involve a definite integral.
- e. What is the value of the limit $\lim_{b \rightarrow \infty} F(b)$? What is its meaning in the context of the problem?

6.5.1 Improper Integrals Involving Unbounded Intervals

In view of the above examples, we see that we may want to integrate over an interval whose upper limit grows without bound. For example, to find the fraction of light bulbs that fail *eventually*, we wish to find

$$\lim_{b \rightarrow \infty} \int_0^b 0.3e^{-0.3t} dt,$$

for which we will also use the notation

$$\int_0^\infty 0.3e^{-0.3t} dt. \quad (6.5.1)$$

Such an integral can be interpreted as the area of an unbounded region, as pictured at right in Figure 6.5.2.

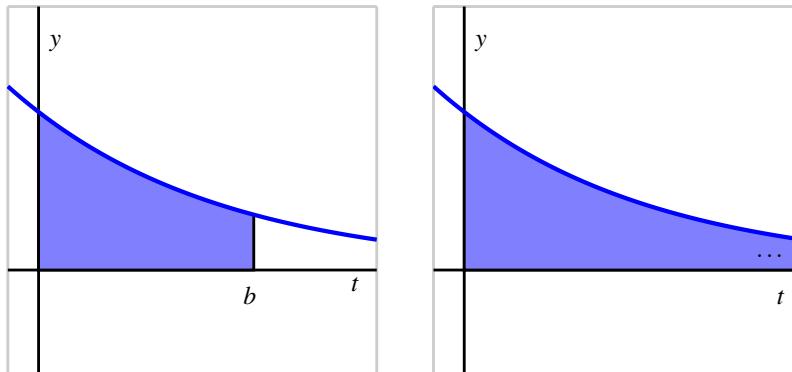


Figure 6.5.2: At left, the area bounded by $p(t) = 0.3e^{-0.3t}$ on the finite interval $[0, b]$; at right, the result of letting $b \rightarrow \infty$. By “...” in the righthand figure, we mean that the region extends to the right without bound.

We call an integral for which the interval of integration is unbounded *improper*. For instance, the integrals

$$\int_1^\infty \frac{1}{x^2} dx, \quad \int_{-\infty}^0 \frac{1}{1+x^2} dx, \quad \text{and} \quad \int_{-\infty}^\infty e^{-x^2} dx$$

are all improper because they have limits of integration that involve ∞ . To evaluate an improper integral we replace it with a limit of proper integrals. That is,

$$\int_0^\infty f(x) dx = \lim_{b \rightarrow \infty} \int_0^b f(x) dx.$$

We first attempt to evaluate $\int_0^b f(x) dx$ using the First FTC, and then evaluate the limit. Is it even possible for the area of an unbounded region to be finite? The following example explores this issue and others in more detail.

Example 6.5.3. In this example we explore the improper integrals $\int_1^\infty \frac{1}{x} dx$ and $\int_1^\infty \frac{1}{x^{3/2}} dx$.

a. First we investigate $\int_1^\infty \frac{1}{x} dx$.

- Use the First FTC to determine the exact values of $\int_1^{10} \frac{1}{x} dx$, $\int_1^{1000} \frac{1}{x} dx$, and $\int_1^{100000} \frac{1}{x} dx$. Then, use your computational device to compute a decimal approximation of each result.
- Use the First FTC to evaluate the definite integral $\int_1^b \frac{1}{x} dx$ (which results in an expression that depends on b).
- Now, use your work from (ii.) to evaluate the limit given by

$$\lim_{b \rightarrow \infty} \int_1^b \frac{1}{x} dx.$$

b. Next, we investigate $\int_1^\infty \frac{1}{x^{3/2}} dx$.

- Use the First FTC to determine the exact values of $\int_1^{10} \frac{1}{x^{3/2}} dx$, $\int_1^{1000} \frac{1}{x^{3/2}} dx$, and $\int_1^{100000} \frac{1}{x^{3/2}} dx$. Then, use your calculator to compute a decimal approximation of each result.
- Use the First FTC to evaluate the definite integral $\int_1^b \frac{1}{x^{3/2}} dx$ (which results in an expression that depends on b).
- Now, use your work from (ii.) to evaluate the limit given by

$$\lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^{3/2}} dx.$$

- Plot the functions $y = \frac{1}{x}$ and $y = \frac{1}{x^{3/2}}$ on the same coordinate axes for the values $x = 0 \dots 10$. How would you compare their behavior as x increases without bound? What is similar? What is different?
- How would you characterize the value of $\int_1^\infty \frac{1}{x} dx$? of $\int_1^\infty \frac{1}{x^{3/2}} dx$? What does this tell us about the respective areas bounded by these two curves for $x \geq 1$?

Hint.

- a. $\frac{1}{x} = x^{-1}$.
- b. $\frac{1}{x^{3/2}} = x^{-3/2}$
- c. Compare how quickly the curves approach the x -axis as $x \rightarrow \infty$.
- d. Remember that $\lim_{b \rightarrow \infty} \frac{1}{b^k} = 0$ as long as $k > 0$.

Answer.

- a. i. $\int_1^{10} \frac{1}{x} dx = \ln(10)$ $\int_1^{1000} \frac{1}{x} dx = \ln(1000)$ $\int_1^{100000} \frac{1}{x} dx = \ln(100000)$
ii. $\int_1^b \frac{1}{x} dx = \ln(b)$.
iii. $\lim_{b \rightarrow \infty} \int_1^b \frac{1}{x} dx = \lim_{b \rightarrow \infty} \ln(b) = \infty$
- b. i. $\int_1^{10} \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{10}}$ $\int_1^{1000} \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{1000}}$ $\int_1^{100000} \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{100000}}$
ii. $\int_1^b \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{b}}$.
iii. $\lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^{3/2}} dx = \lim_{b \rightarrow \infty} \left(2 - \frac{2}{\sqrt{b}}\right) = 2$
- c. Both graphs have a vertical asymptote at $x = 0$ and for both graphs, the x -axis is a horizontal asymptote. However, the graph of $y = \frac{1}{x^{3/2}}$ will "approach the x -axis faster" than the graph of $y = \frac{1}{x}$.
- d. The area bounded by the graph of $y = \frac{1}{x}$, the x -axis, and the vertical line $x = 1$ is infinite or unbounded. However, The area bounded by the graph of $y = \frac{1}{x^{3/2}}$, the x -axis, and the vertical line $x = 1$ is equal to 2.

Solution.

- a. i. $\int_1^{10} \frac{1}{x} dx = \ln(10)$ $\int_1^{1000} \frac{1}{x} dx = \ln(1000)$ $\int_1^{100000} \frac{1}{x} dx = \ln(100000)$
ii. $\int_1^b \frac{1}{x} dx = \ln(b)$.
iii. $\lim_{b \rightarrow \infty} \int_1^b \frac{1}{x} dx = \lim_{b \rightarrow \infty} \ln(b) = \infty$
- b. i. $\int_1^{10} \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{10}}$ $\int_1^{1000} \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{1000}}$ $\int_1^{100000} \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{100000}}$
ii. $\int_1^b \frac{1}{x^{3/2}} dx = 2 - \frac{2}{\sqrt{b}}$.
iii. $\lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^{3/2}} dx = \lim_{b \rightarrow \infty} \left(2 - \frac{2}{\sqrt{b}}\right) = 2$
- c. Both graphs have a vertical asymptote at $x = 0$ and for both graphs, the x -axis is a horizontal asymptote. However, the graph of $y = \frac{1}{x^{3/2}}$ will "approach the x -axis faster" than the graph of $y = \frac{1}{x}$.
- d. The area bounded by the graph of $y = \frac{1}{x}$, the x -axis, and the vertical line $x = 1$ is infinite or unbounded. However, The area bounded by the graph of $y = \frac{1}{x^{3/2}}$, the x -axis, and the vertical line $x = 1$ is equal to 2.

6.5.2 Convergence and Divergence

Example 6.5.3 suggests that $\lim_{b \rightarrow \infty} \int_1^b f(x) dx$ is either finite or infinite (or it doesn't exist). With these possibilities in mind, we introduce the following terminology.

If $f(x)$ is nonnegative for $x \geq a$, then we say that the improper integral $\int_a^\infty f(x) dx$ **converges** provided that

$$\lim_{b \rightarrow \infty} \int_a^b f(x) dx$$

exists and is finite. Otherwise, we say that $\int_a^\infty f(x) dx$ **diverges**.

We will restrict our interest to improper integrals for which the integrand is nonnegative. Also, we require that $\lim_{x \rightarrow \infty} f(x) = 0$, for if f does not approach 0 as $x \rightarrow \infty$, then it is impossible for $\int_a^\infty f(x) dx$ to converge.

Example 6.5.4. Determine whether each of the following improper integrals converges or diverges. For each integral that converges, find its exact value.

a. $\int_1^\infty \frac{1}{x^2} dx$

d. $\int_4^\infty \frac{3}{(x+2)^{5/4}} dx$

b. $\int_0^\infty e^{-x/4} dx$

e. $\int_0^\infty x e^{-x/4} dx$

c. $\int_2^\infty \frac{9}{(x+5)^{2/3}} dx$

f. $\int_1^\infty \frac{1}{x^p} dx$, where p is a positive real number

Hint.

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

b. Recall that for $k > 0$, $e^{-kt} \rightarrow 0$ as $t \rightarrow \infty$.

c. $\frac{9}{(x+5)^{2/3}} = 9(x+5)^{-2/3}$.

d. Compare (c).

e. Compare (b), after integrating by parts.

f. Try two cases: $0 < p < 1$ and $1 < p$.

Answer.

a. $\int_1^\infty \frac{1}{x^2} dx = 1$

b. $\int_0^\infty e^{-x/4} dx = 4$

c. $\int_2^\infty \frac{9}{(x+5)^{2/3}} dx = \infty$

d. $\int_4^\infty \frac{3}{(x+2)^{5/4}} dx = \frac{12}{6^{1/4}}$

e. $\int_0^\infty xe^{-x/4} dx = 16$

f. If $0 < p < 1$, $\int_1^\infty \frac{1}{x^p} dx$ diverges, while if $p > 1$, the integral converges.

Solution.

a.

$$\begin{aligned}\int_1^\infty \frac{1}{x^2} dx &= \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2} dx \\ &= \lim_{b \rightarrow \infty} -\frac{1}{x} \Big|_1^b \\ &= \lim_{b \rightarrow \infty} \left(-\frac{1}{b} + 1 \right) \\ &= 1\end{aligned}$$

b.

$$\begin{aligned}\int_0^\infty e^{-x/4} dx &= \lim_{b \rightarrow \infty} \int_0^b e^{-x/4} dx \\ &= \lim_{b \rightarrow \infty} (-4e^{-x/4}) \Big|_0^b \\ &= \lim_{b \rightarrow \infty} (-4e^{-b/4} + 4) \\ &= 4\end{aligned}$$

c.

$$\begin{aligned}\int_2^\infty \frac{9}{(x+5)^{2/3}} dx &= \lim_{b \rightarrow \infty} \int_2^b \frac{9}{(x+5)^{2/3}} dx \\ &= \lim_{b \rightarrow \infty} 27(x+5)^{1/3} \Big|_2^b \\ &= \lim_{b \rightarrow \infty} (27(b+5)^{1/3} - 27(7)^{1/3}) \\ &= \infty\end{aligned}$$

d.

$$\begin{aligned}\int_4^\infty \frac{3}{(x+2)^{5/4}} dx &= \lim_{b \rightarrow \infty} \int_4^b \frac{3}{(x+2)^{5/4}} dx \\ &= \lim_{b \rightarrow \infty} \left(-\frac{12}{(b+2)^{1/4}} + \frac{12}{6^{1/4}} \right) \\ &= \frac{12}{6^{1/4}}\end{aligned}$$

e.

$$\begin{aligned}
 \int_0^\infty xe^{-x/4} dx &= \lim_{b \rightarrow \infty} \int_0^b xe^{-x/4} dx \\
 &= \lim_{b \rightarrow \infty} (-4xe^{-x/4} - 16e^{-x/4}) \Big|_0^b \\
 &= 16
 \end{aligned}$$

f. We assume that p is a positive real number.

$$\begin{aligned}
 \int_1^\infty \frac{1}{x^p} dx &= \lim_{b \rightarrow \infty} \int_1^b x^{-p} dx \\
 &= \lim_{b \rightarrow \infty} \left(-\frac{b^{1-p}}{p-1} + \frac{1}{p-1} \right)
 \end{aligned}$$

We now use two cases:

- (Assume $0 < p < 1$.) In this case, $1 - p > 0$ and so $\lim_{b \rightarrow \infty} b^{1-p} = \infty$ and

$$\lim_{b \rightarrow \infty} \left(-\frac{b^{1-p}}{p-1} + \frac{1}{p-1} \right) = \infty$$

Using this and equation (1), we see that $\int_1^\infty \frac{1}{x^p} dx$ diverges.

- (Assume $p > 1$.) In this case, $1 - p < 0$ and so $\lim_{b \rightarrow \infty} b^{1-p} = 0$ and

$$\lim_{b \rightarrow \infty} \left(-\frac{b^{1-p}}{p-1} + \frac{1}{p-1} \right) = \frac{1}{p-1}$$

Using this and equation (1), we see that $\int_1^\infty \frac{1}{x^p} dx$ converges to $\frac{1}{p-1}$.

6.5.3 Improper Integrals Involving Unbounded Integrands

An integral is also called improper if the integrand is unbounded on the interval of integration. For example, consider

$$\int_0^1 \frac{1}{\sqrt{x}} dx.$$

Because $f(x) = \frac{1}{\sqrt{x}}$ has a vertical asymptote at $x = 0$, f is not continuous on $[0, 1]$, and the integral represents the area of the unbounded region shown at right in Figure 6.5.5.

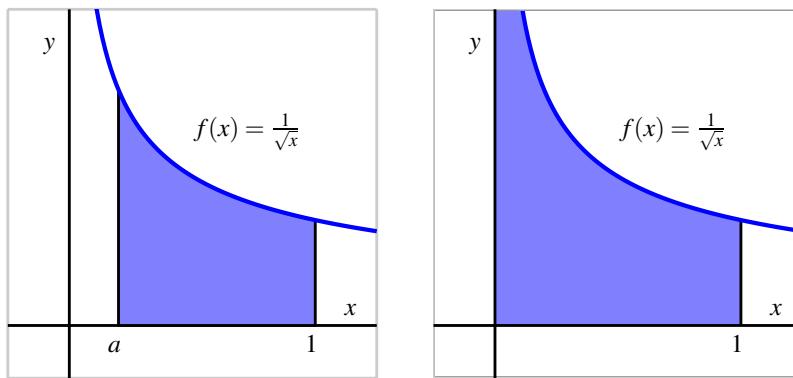


Figure 6.5.5: At left, the area bounded by $f(x) = \frac{1}{\sqrt{x}}$ on the finite interval $[a, 1]$; at right, the result of letting $a \rightarrow 0^+$, where we see that the shaded region will extend vertically without bound.

We address the problem of the integrand being unbounded by replacing the improper integral with a limit of proper integrals. For example, to evaluate $\int_0^1 \frac{1}{\sqrt{x}} dx$, we replace 0 with a and let a approach 0 from the right. Thus,

$$\int_0^1 \frac{1}{\sqrt{x}} dx = \lim_{a \rightarrow 0^+} \int_a^1 \frac{1}{\sqrt{x}} dx.$$

We evaluate the proper integral $\int_a^1 \frac{1}{\sqrt{x}} dx$, and then take the limit. We will say that the improper integral converges if this limit exists, and diverges otherwise. In this example, we observe that

$$\begin{aligned} \int_0^1 \frac{1}{\sqrt{x}} dx &= \lim_{a \rightarrow 0^+} \int_a^1 \frac{1}{\sqrt{x}} dx \\ &= \lim_{a \rightarrow 0^+} 2\sqrt{x} \Big|_a^1 \\ &= \lim_{a \rightarrow 0^+} 2\sqrt{1} - 2\sqrt{a} \\ &= 2, \end{aligned}$$

so the improper integral $\int_0^1 \frac{1}{\sqrt{x}} dx$ converges (to the value 2).

We have to be particularly careful with unbounded integrands, for they may arise in ways that may not initially be obvious. Consider, for instance, the integral

$$\int_1^3 \frac{1}{(x-2)^2} dx.$$

At first glance we might think that we can simply apply the Fundamental Theorem of Calculus by antiderivativing $\frac{1}{(x-2)^2}$ to get $-\frac{1}{x-2}$ and then evaluating from 1 to 3. Were we to do

so, we would be erroneously applying the FTC because $f(x) = \frac{1}{(x-2)^2}$ fails to be continuous throughout the interval, as seen in Figure 6.5.6.

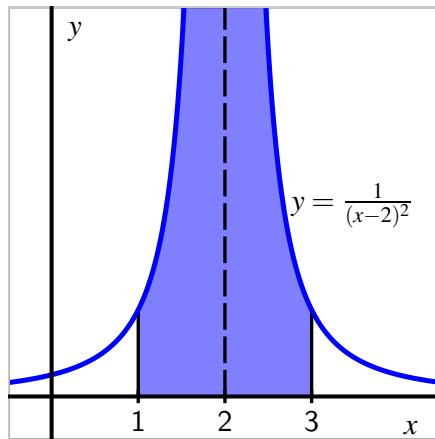


Figure 6.5.6: The function $f(x) = \frac{1}{(x-2)^2}$ on an interval including $x = 2$.

Such an incorrect application of the FTC leads to an impossible result (-2), which would itself suggest that something we did must be wrong. Instead, we must address the vertical asymptote at $x = 2$ by writing

$$\int_1^3 \frac{1}{(x-2)^2} dx = \lim_{a \rightarrow 2^-} \int_1^a \frac{1}{(x-2)^2} dx + \lim_{b \rightarrow 2^+} \int_b^3 \frac{1}{(x-2)^2} dx.$$

We then evaluate two separate limits of proper integrals. For instance, doing so for the integral with a approaching 2 from the left, we find

$$\begin{aligned} \int_1^2 \frac{1}{(x-2)^2} dx &= \lim_{a \rightarrow 2^-} \int_1^a \frac{1}{(x-2)^2} dx \\ &= \lim_{a \rightarrow 2^-} -\frac{1}{(x-2)} \Big|_1^a \\ &= \lim_{a \rightarrow 2^-} -\frac{1}{(a-2)} + \frac{1}{1-2} \\ &= \infty, \end{aligned}$$

since $\frac{1}{a-2} \rightarrow -\infty$ as a approaches 2 from the left. Thus, the improper integral $\int_1^2 \frac{1}{(x-2)^2} dx$ diverges; similar work shows that $\int_2^3 \frac{1}{(x-2)^2} dx$ also diverges. From either of these two results, we can conclude that the original integral, $\int_1^3 \frac{1}{(x-2)^2} dx$ diverges, too.

Example 6.5.7. For each of the following definite integrals, decide whether the integral is improper or not. If the integral is proper, evaluate it using the First FTC. If the integral is improper, determine whether or not the integral converges or diverges; if the integral converges, find its exact value.

a. $\int_0^1 \frac{1}{x^{1/3}} dx$

d. $\int_{-2}^2 \frac{1}{x^2} dx$

b. $\int_0^2 e^{-x} dx$

e. $\int_0^{\pi/2} \tan(x) dx$

c. $\int_1^4 \frac{1}{\sqrt{4-x}} dx$

f. $\int_0^1 \frac{1}{\sqrt{1-x^2}} dx$

Hint.a. $\frac{1}{x^{1/3}}$ is undefined at $x = 0$.

b. This integral is a proper one.

c. $\frac{1}{\sqrt{4-x}}$ is undefined at $x = 4$.d. Be careful about what happens at $x = 0$; split the original integral into two integrals that involve $x = 0$.e. $\tan(x)$ is undefined at $x = \pi/2$.f. Recall that you know an antiderivative for $\frac{1}{\sqrt{1-x^2}}$.**Answer.**

a. $\int_0^1 \frac{1}{x^{1/3}} dx = \frac{3}{2}$

b. $\int_0^2 e^{-x} dx = 1 - e^{-2}$

c. $\int_0^4 \frac{1}{\sqrt{4-x}} dx = 4$

d. $\int_{-2}^2 \frac{1}{x^2} dx$ diverges.

e. $\int_0^{\pi/2} \tan(x) dx = \infty$

f. $\int_0^1 \frac{1}{\sqrt{1-x^2}} dx = \frac{\pi}{2}$

Solution.

a.

$$\begin{aligned}
 \int_0^1 \frac{1}{x^{1/3}} dx &= \lim_{a \rightarrow 0^+} \int_0^1 \frac{1}{x^{1/3}} dx \\
 &= \lim_{a \rightarrow 0^+} \left(\frac{3}{2} - \frac{3}{2} a^{2/3} \right) \\
 &= \frac{3}{2}
 \end{aligned}$$

b.

$$\begin{aligned}\int_0^2 e^{-x} dx &= -e^{-x} \Big|_0^2 \\ &= 1 - e^{-2}\end{aligned}$$

c.

$$\begin{aligned}\int_0^4 \frac{1}{\sqrt{4-x}} dx &= \lim_{b \rightarrow 4^-} \int_0^b \frac{1}{\sqrt{4-x}} dx \\ &= \lim_{b \rightarrow 4^-} -2\sqrt{4-x} \Big|_0^b \\ &= 4\end{aligned}$$

d.

$$\int_{-2}^2 \frac{1}{x^2} dx = \int_{-2}^0 \frac{1}{x^2} dx + \int_0^2 \frac{1}{x^2} dx$$

However, each of the improper integrals on the right side of the equation diverges. For example,

$$\begin{aligned}\int_0^2 \frac{1}{x^2} dx &= \lim_{a \rightarrow 0^+} \int_a^2 \frac{1}{x^2} dx \\ &= \lim_{a \rightarrow 0^+} \left(-\frac{1}{2} + \frac{1}{a} \right) \\ &= \infty\end{aligned}$$

So $\int_{-2}^2 \frac{1}{x^2} dx$ diverges.

e.

$$\begin{aligned}\int_0^{\pi/2} \tan(x) dx &= \lim_{b \rightarrow \frac{\pi}{2}^+} \int_0^b \tan(x) dx \\ &= \lim_{b \rightarrow \frac{\pi}{2}^+} (-\ln(\cos(b)) + \ln(\cos(0))) \\ &= \infty\end{aligned}$$

f.

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{1-x^2}} dx &= \lim_{b \rightarrow 1^+} \int_0^b \frac{1}{\sqrt{1-x^2}} dx \\ &= \lim_{b \rightarrow 1^+} (\arcsin(b) - \arcsin(0)) \\ &= \frac{\pi}{2}\end{aligned}$$

6.5.4 Summary

- An integral $\int_a^b f(x) dx$ can be improper if at least one of a or b is $\pm\infty$, making the interval unbounded, or if f has a vertical asymptote at $x = c$ for some value of c that satisfies $a \leq c \leq b$. One reason that improper integrals are important is that certain probabilities can be represented by integrals that involve infinite limits.
- When we encounter an improper integral, we work to understand it by replacing the improper integral with a limit of proper integrals. For instance, we write

$$\int_a^{\infty} f(x) dx = \lim_{b \rightarrow \infty} \int_a^b f(x) dx,$$

and then work to determine whether the limit exists and is finite. For any improper integral, if the resulting limit of proper integrals exists and is finite, we say the improper integral converges. Otherwise, the improper integral diverges.

- An important class of improper integrals is given by

$$\int_1^{\infty} \frac{1}{x^p} dx$$

where p is a positive real number. We can show that this improper integral converges whenever $p > 1$, and diverges whenever $0 < p \leq 1$. A related class of improper integrals is $\int_0^1 \frac{1}{x^p} dx$, which converges for $0 < p < 1$, and diverges for $p \geq 1$.

6.5.5 Exercises

1. (An improper integral on a finite interval) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (An improper integral on an infinite interval) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (An improper integral involving a ratio of exponential functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (A subtle improper integral) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (An improper integral involving a ratio of trigonometric functions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. Determine, with justification, whether each of the following improper integrals converges or diverges.
 - $\int_e^{\infty} \frac{\ln(x)}{x} dx$
 - $\int_e^{\infty} \frac{1}{x \ln(x)} dx$
 - $\int_e^{\infty} \frac{1}{x(\ln(x))^2} dx$
 - $\int_e^{\infty} \frac{1}{x(\ln(x))^p} dx$, where p is a positive real number

e. $\int_0^1 \frac{\ln(x)}{x} dx$
 f. $\int_0^1 \ln(x) dx$

Answer.

- Diverges.
- Diverges.
- Converges to 1.
- $\int_e^\infty \frac{1}{x(\ln(x))^p} dx$ diverges if $p \leq 1$ and converges to $\frac{1}{p-1}$ if $p > 1$.
- Diverges.
- Converges to -1 .

Solution.

- We write the improper integral as a limit and use the substitution $u = \ln(x)$ and $du = \frac{1}{x} dx$ to evaluate the definite integral, then take the limit:

$$\begin{aligned} \int_e^\infty \frac{\ln(x)}{x} dx &= \lim_{b \rightarrow \infty} \int_e^b \frac{\ln(x)}{x} dx \\ &= \lim_{b \rightarrow \infty} \int_1^{\ln(b)} u du \\ &= \lim_{b \rightarrow \infty} \left(\frac{u^2}{2} \right) \Big|_1^{\ln(b)} \\ &= \lim_{b \rightarrow \infty} \left(\frac{(\ln(b))^2}{2} - \frac{1}{2} \right) \\ &= \infty. \end{aligned}$$

So the improper integral $\int_e^\infty \frac{\ln(x)}{x} dx$ diverges.

- We write the improper integral as a limit and use the substitution $u = \ln(x)$ and $du = \frac{1}{x} dx$ to reevaluate the definite integral, then take the limit.:

$$\begin{aligned} \int_e^\infty \frac{1}{x \ln(x)} dx &= \lim_{b \rightarrow \infty} \int_e^b \frac{1}{x \ln(x)} dx \\ &= \lim_{b \rightarrow \infty} \int_1^{\ln(b)} \frac{1}{u} du \\ &= \lim_{b \rightarrow \infty} (\ln(|u|)) \Big|_1^{\ln(b)} \\ &= \lim_{b \rightarrow \infty} (\ln(\ln(|b|)) - \ln(1)) \\ &= \infty. \end{aligned}$$

So the improper integral $\int_e^\infty \frac{1}{x \ln(x)} dx$ diverges.

- We again write the improper integral as a limit and use the substitution $u = \ln(x)$ and $du = \frac{1}{x} dx$ to evaluate the definite integral, then take the limit:

$$\int_e^\infty \frac{1}{x(\ln(x))^2} dx = \lim_{b \rightarrow \infty} \int_e^b \frac{1}{x(\ln(x))^2} dx$$

$$\begin{aligned}
&= \lim_{b \rightarrow \infty} \int_1^{\ln(b)} \frac{1}{u^2} du \\
&= \lim_{b \rightarrow \infty} (-u^{-1}) \Big|_1^{\ln(b)} \\
&= \lim_{b \rightarrow \infty} \left(-\frac{1}{\ln(b)} - (-1) \right) \\
&= 1.
\end{aligned}$$

So the improper integral $\int_e^\infty \frac{1}{x(\ln(x))^2} dx$ converges to 1.

- d. We again write the improper integral as a limit and use the substitution $u = \ln(x)$ and $du = \frac{1}{x} dx$ to rewrite the definite integral. If $p = 1$, then the integral diverges by part (b). If $p \neq 1$, we have

$$\begin{aligned}
\int_e^\infty \frac{1}{x(\ln(x))^p} dx &= \lim_{b \rightarrow \infty} \int_e^b \frac{1}{x(\ln(x))^p} dx \\
&= \lim_{b \rightarrow \infty} \int_1^{\ln(b)} \frac{1}{u^p} du \\
&= \lim_{b \rightarrow \infty} \int_1^{\ln(b)} u^{-p} du \\
&= \lim_{b \rightarrow \infty} \left(\frac{u^{-p+1}}{-p+1} \right) \Big|_1^{\ln(b)} \\
&= \frac{1}{1-p} \lim_{b \rightarrow \infty} ((\ln(b))^{1-p} - 1) \\
&= 1.
\end{aligned}$$

If $p < 1$, then $1-p > 0$ and

$$\frac{1}{1-p} \lim_{b \rightarrow \infty} ((\ln(b))^{1-p} - 1) = \left(\frac{1}{1-p} \right) \infty = \infty.$$

In this case the improper integral $\int_e^\infty \frac{1}{x(\ln(x))^p} dx$ diverges.

If $p > 1$, then $1-p < 0$ and

$$\frac{1}{1-p} \lim_{b \rightarrow \infty} ((\ln(b))^{1-p} - 1) = \frac{1}{1-p} \lim_{b \rightarrow \infty} \left(\frac{1}{(\ln(b))^{p-1}} - 1 \right) = \left(\frac{1}{1-p} \right) (0-1) = \frac{1}{p-1}$$

In this case the improper integral $\int_e^\infty \frac{1}{x(\ln(x))^p} dx$ converges to $\frac{1}{p-1}$.

Thus, the improper integral $\int_e^\infty \frac{1}{x(\ln(x))^p} dx$ diverges if $p \leq 1$ and converges to $\frac{1}{p-1}$ if $p > 1$.

- e. This integral is improper since the integrand is undefined at $x = 0$. So we write the improper integral as a limit, use the substitution $u = \ln(x)$ and $du = \frac{1}{x} dx$ to evaluate the definite integral, then take the limit:

$$\int_0^1 \frac{\ln(x)}{x} dx = \lim_{a \rightarrow 0^+} \int_a^1 \frac{\ln(x)}{x} dx$$

$$\begin{aligned}
&= \lim_{a \rightarrow 0^+} \int_{\ln(a)}^0 u \, du \\
&= \lim_{a \rightarrow 0^+} \left(\frac{u^2}{2} \right) \Big|_{\ln(a)}^0 \\
&= \lim_{a \rightarrow 0^+} \left(-\frac{(\ln(a))^2}{2} \right) \\
&= -\infty.
\end{aligned}$$

So the improper integral $\int_0^1 \frac{\ln(x)}{x} dx$ diverges.

- f. This integral is improper since the integrand is undefined at $x = 0$. So we write the improper integral as a limit, then use parts with $u = \ln(x)$, $du = \frac{1}{x} dx$, $dv = dx$, and $v = x$ to evaluate the definite integral, then take the limit:

$$\begin{aligned}
\int_0^1 \ln(x) \, dx &= \lim_{a \rightarrow 0^+} \int_a^1 \ln(x) \, dx \\
&= \lim_{a \rightarrow 0^+} \left(x \ln(x) \Big|_a^1 - \int_a^1 dx \right) \\
&= \lim_{a \rightarrow 0^+} \left(-a \ln(a) - x \Big|_{\ln(a)}^0 \right) \\
&= \lim_{a \rightarrow 0^+} (-a \ln(a) - (1 - a)) \\
&= \lim_{a \rightarrow 0^+} (-a \ln(a)) - 1.
\end{aligned}$$

The limit $\lim_{a \rightarrow 0^+} (-a \ln(a))$ has the form $0 \times \infty$, which is indeterminate, so we need to rewrite it in a form in which we can apply L'Hopital's Rule. Now,

$$\lim_{a \rightarrow 0^+} (-a \ln(a)) = \lim_{a \rightarrow 0^+} \left(\frac{-\ln(a)}{\frac{1}{a}} \right),$$

and this latter limit is of the form $\frac{\infty}{\infty}$. Applying L'Hopital's Rule gives us

$$\begin{aligned}
\lim_{a \rightarrow 0^+} \left(\frac{-\ln(a)}{\frac{1}{a}} \right) &= \lim_{a \rightarrow 0^+} \left(\frac{-\frac{1}{a}}{-\frac{1}{a^2}} \right) \\
&= \lim_{a \rightarrow 0^+} a \\
&= 0
\end{aligned}$$

Returning to our earlier established equality that

$$\int_0^1 \ln(x) \, dx = \lim_{a \rightarrow 0^+} (-a \ln(a)) - 1,$$

this most recent limit shows that the improper integral $\int_0^1 \ln(x) \, dx$ converges to -1 .

7. Sometimes we may encounter an improper integral for which we cannot easily evaluate the limit of the corresponding proper integrals. For instance, consider $\int_1^\infty \frac{1}{1+x^3} dx$. While

it is hard (or perhaps impossible) to find an antiderivative for $\frac{1}{1+x^3}$, we can still determine whether or not the improper integral converges or diverges by comparison to a simpler one. Observe that for all $x > 0$, $1+x^3 > x^3$, and therefore

$$\frac{1}{1+x^3} < \frac{1}{x^3}.$$

It therefore follows that

$$\int_1^b \frac{1}{1+x^3} dx < \int_1^b \frac{1}{x^3} dx$$

for every $b > 1$. If we let $b \rightarrow \infty$ so as to consider the two improper integrals $\int_1^\infty \frac{1}{1+x^3} dx$ and $\int_1^\infty \frac{1}{x^3} dx$, we know that the larger of the two improper integrals converges. And thus, since the smaller one lies below a convergent integral, it follows that the smaller one must converge, too. In particular, $\int_1^\infty \frac{1}{1+x^3} dx$ must converge, even though we never explicitly evaluated the corresponding limit of proper integrals. We use this idea and similar ones in the exercises that follow.

- Explain why $x^2 + x + 1 > x^2$ for all $x \geq 1$, and hence show that $\int_1^\infty \frac{1}{x^2+x+1} dx$ converges by comparison to $\int_1^\infty \frac{1}{x^2} dx$.
- Observe that for each $x > 1$, $\ln(x) < x$. Explain why

$$\int_2^b \frac{1}{x} dx < \int_2^b \frac{1}{\ln(x)} dx$$

for each $b > 2$. Why must it be true that $\int_2^b \frac{1}{\ln(x)} dx$ diverges?

- Explain why $\sqrt{\frac{x^4+1}{x^4}} > 1$ for all $x > 1$. Then, determine whether or not the improper integral

$$\int_1^\infty \frac{1}{x} \cdot \sqrt{\frac{x^4+1}{x^4}} dx$$

converges or diverges.

Answer.

- converges
- diverges
- diverges

Solution.

- Since $x \geq 1$, it follows that $x+1 \geq 1 > 0$, so adding x^2 to both sides of this inequality shows that

$$x^2 + x + 1 > x^2$$

for all $x \geq 1$. Taking reciprocals (which reverses the inequality),

$$\frac{1}{x^2 + x + 1} < \frac{1}{x^2}$$

and thus for any $b > 1$,

$$\int_1^b \frac{1}{x^2 + x + 1} dx < \int_1^b \frac{1}{x^2} dx.$$

Taking the limit as $b \rightarrow \infty$, it follows that

$$\int_1^\infty \frac{1}{x^2 + x + 1} dx < \int_1^\infty \frac{1}{x^2} dx,$$

and since we know that $\int_1^\infty \frac{1}{x^2} dx$ converges, we conclude that $\int_1^\infty \frac{1}{x^2 + x + 1} dx$ also converges.

- b. Since for each $x > 1$, $\ln(x) < x$, it also holds that $\ln(x) < x$ for all $x > 2$. Taking reciprocals,

$$\frac{1}{x} < \frac{1}{\ln(x)}$$

and thus

$$\int_2^b \frac{1}{x} dx < \int_2^b \frac{1}{\ln(x)} dx$$

for each $b > 2$. Taking the limit as $b \rightarrow \infty$, we have

$$\int_2^\infty \frac{1}{x} dx \leq \int_2^\infty \frac{1}{\ln(x)} dx$$

and since we know that $\int_1^\infty \frac{1}{x} dx$ diverges, it follows that $\int_2^\infty \frac{1}{\ln(x)} dx$ also diverges.

- c. Since $x^4 + 1 > x^4$ for all x , it follows that

$$\frac{x^4 + 1}{x^4} > 1$$

for all $x > 1$, and thus taking square roots,

$$\sqrt{\frac{x^4 + 1}{x^4}} > 1$$

for all $x > 1$. Integrating from 1 to b and then letting $b \rightarrow \infty$, we can next say that

$$\int_1^\infty \sqrt{\frac{x^4 + 1}{x^4}} dx \geq \int_1^\infty 1 dx$$

It is obvious that $\int_1^\infty 1 dx$ diverges since the integrand does not tend to 0. By the last inequality above, we can conclude that

$$\int_1^\infty \frac{1}{x} \cdot \sqrt{\frac{x^4 + 1}{x^4}} dx$$

diverges.

Chapter 7

Differential Equations

7.1 An Introduction to Differential Equations

Motivating Questions

- This chapter needs learning objectives
-

7.1.1 What is a differential equation?

Differential equations arise frequently in our every day world. For instance, you may hear a bank advertising:

Your money will grow at a 3% annual interest rate with us.

This innocuous statement is really a differential equation. Let's translate: $A(t)$ will be amount of money you have in your account at time t . The rate at which your money grows is the derivative dA/dt , and we are told that this rate is $0.03A$. This leads to the differential equation

$$\frac{dA}{dt} = 0.03A.$$

This differential equation has a slightly different feel than the previous equation $\frac{ds}{dt} = 4t + 1$. In the earlier example, the rate of change depends only on the independent variable t , and we may find $s(t)$ by integrating the velocity $4t + 1$. In the banking example, however, the rate of change depends on the dependent variable A , so we'll need some new techniques in order to find $A(t)$.

Example 7.1.1. Express the following statements as differential equations. In each case, you will need to introduce notation to describe the important quantities in the statement so be sure to clearly state what your notation means.

- The population of a town grows continuously at an annual rate of 1.25%.
- A radioactive sample loses 5.6% of its mass every day.
- You have a bank account that continuously earns 4% interest every year. At the same time, you withdraw money continually from the account at the rate of \$1000 per year.
- A cup of hot chocolate is sitting in a 70° room. The temperature of the hot chocolate cools continuously by 10% of the difference between the hot chocolate's temperature and the room temperature every minute.
- A can of cold soda is sitting in a 70° room. The temperature of the soda warms continuously at the rate of 10% of the difference between the soda's temperature and the room's temperature every minute.

Hint.

- Small hints for each of the prompts above.

Answer.

- Let P be the population t the time in years; $\frac{dP}{dt} = 0.0125P$.
- Let m be the mass t the time in days; $\frac{dm}{dt} = -0.056m$.
- Let B be the balance t be time in years; $\frac{dB}{dt} = 0.04B - 1000$.
- Let t be time in minutes H the temperature of the hot chocolate; $\frac{dH}{dt} = -0.1(H - 70)$.
- Let t be time in minutes and H the temperature of the soda;

$$\frac{dH}{dt} = 0.1(70 - H) = -0.1(H - 70).$$

Solution.

- Let P be the population of the town and let t be the time in years. The population of the town grows continuously at an annual rate of 1.25% means that

$$\frac{dP}{dt} = 0.0125P.$$

- Let m be the mass of a radioactive sample and let t be the time in days. The sample loses 5.6% of its mass every day means that

$$\frac{dm}{dt} = -0.056m.$$

- c. Let B be the balance (in dollars) in the account and let t be the time in years. The money in the account continuously earns 4% interest every year and \$1000 is withdrawn every year means that

$$\frac{dB}{dt} = 0.04B - 1000.$$

- d. The temperature of a room is 70° . Let t be time in minutes. The temperature H of a cup of hot chocolate cools continuously by 10% of the difference between the hot chocolate's temperature and the room temperature every minute means that

$$\frac{dH}{dt} = -0.1(H - 70).$$

- e. The temperature of a room is 70° . Let t be time in minutes. The temperature H of a can of soda warms continuously by 10% of the difference between the soda's temperature and the room temperature every minute means that

$$\frac{dH}{dt} = 0.1(70 - H) = -0.1(H - 70).$$

7.1.2 Differential equations in the world around us

Differential equations give a natural way to describe phenomena we see in the real world. For instance, physical principles are frequently expressed as a description of how a quantity changes. A good example is Newton's Second Law, which says:

The product of an object's mass and acceleration equals the force applied to it.

For instance, when gravity acts on an object near the earth's surface, it exerts a force equal to mg , the mass of the object times the gravitational constant g . We therefore have

$$ma = mg, \text{ or}$$

$$\frac{dv}{dt} = g,$$

where v is the velocity of the object, and $g = 9.8$ meters per second squared. Notice that this physical principle does not tell us what the object's velocity is, but rather how the object's velocity changes.

Example 7.1.2. Shown below are two graphs depicting the velocity of falling objects. On the left is the velocity of a skydiver, while on the right is the velocity of a meteorite entering the Earth's atmosphere.

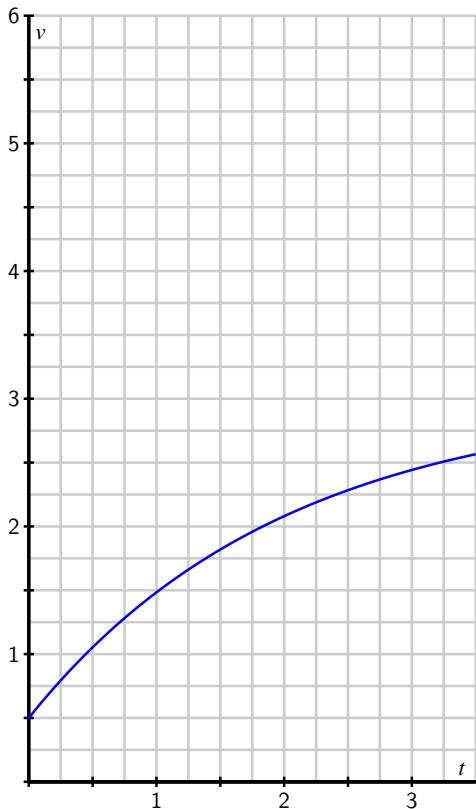


Figure 7.1.3: A skydiver's velocity.

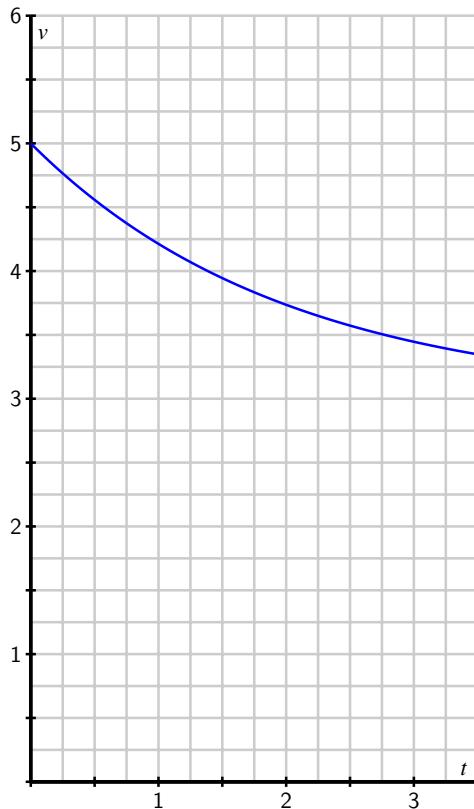
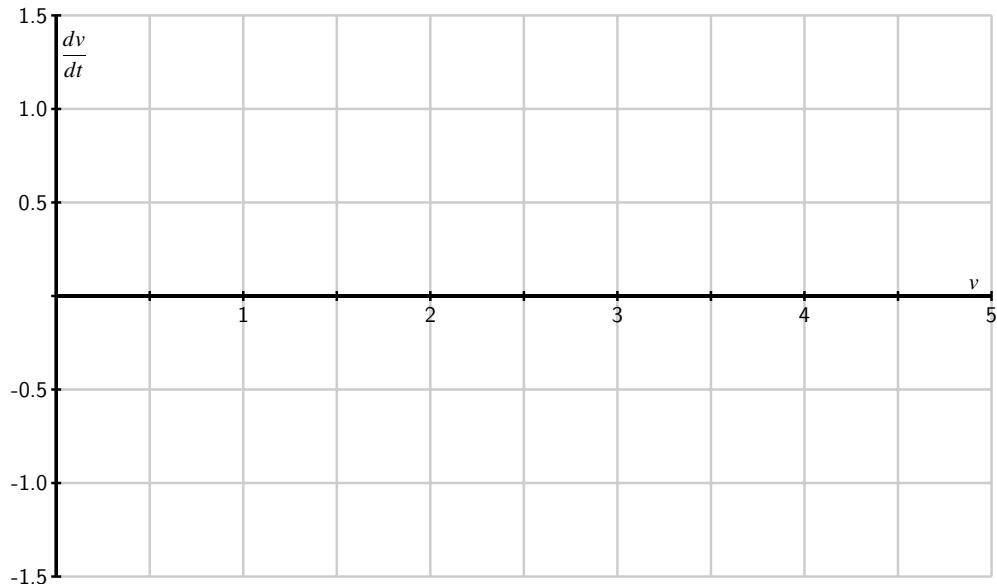


Figure 7.1.4: A meteorite's velocity.

- a. Begin with the skydiver's velocity and use the given graph to measure the rate of change dv/dt when the velocity is $v = 0.5, 1.0, 1.5, 2.0$, and 2.5 . Plot your values on the graph below. You will want to think carefully about this: you are plotting the derivative dv/dt as a function of *velocity*.



- b. Now do the same thing with the meteorite's velocity: use the given graph to measure the rate of change dv/dt when the velocity is $v = 3.5, 4.0, 4.5$, and 5.0 . Plot your values on the graph above.
- c. You should find that all your points lie on a line. Write the equation of this line being careful to use proper notation for the quantities on the horizontal and vertical axes.
- d. The relationship you just found is a differential equation. Write a complete sentence that explains its meaning.
- e. By looking at the differential equation, determine the values of the velocity for which the velocity increases.
- f. By looking at the differential equation, determine the values of the velocity for which the velocity decreases.
- g. By looking at the differential equation, determine the values of the velocity for which the velocity remains constant.

Hint.

- a. Hints for each of the prompts above.

Answer.

- a. For the skydiver:

$$\left. \frac{dv}{dt} \right|_{(v=0.5)} \approx 1.5$$

$$\left. \frac{dv}{dt} \right|_{(v=1)} \approx 1.2$$

$$\left. \frac{dv}{dt} \right|_{(v=1.5)} \approx 0.9$$

$$\frac{dv}{dt} \Big|_{(v=2)} \approx 0.6$$

$$\frac{dv}{dt} \Big|_{(v=2.5)} \approx 0.3$$

b. For the meteorite:

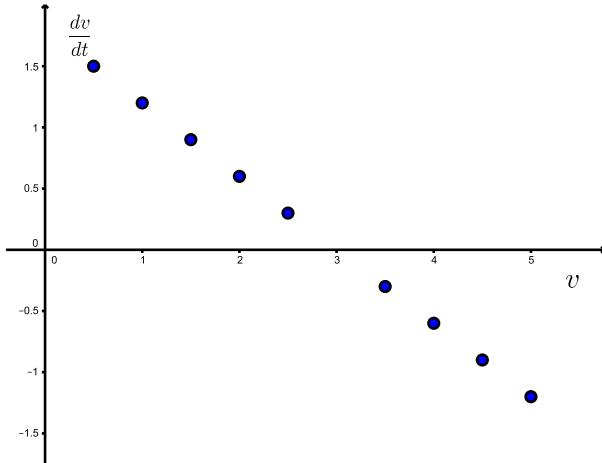
$$\frac{dv}{dt} \Big|_{(v=3.5)} \approx -0.3$$

$$\frac{dv}{dt} \Big|_{(v=4)} \approx -0.9$$

$$\frac{dv}{dt} \Big|_{(v=4)} \approx -0.6$$

$$\frac{dv}{dt} \Big|_{(v=5)} \approx -1.2$$

A graph of the points from parts (a) and (b) is shown in the following diagram:



c. $\frac{dv}{dt} = -0.6v + 1.8$.

d. The rate of change of velocity with respect to time is a linear function of velocity.

e. $0 < v < 3$.

f. $3 < v < 5$.

g. $v = 3$.

Solution. Each graph is a graph of velocity v as a function of time t . To estimate $\frac{dv}{dt}$ at a particular time, we must estimate the slope of the tangent line for the graph of v at that point. For parts (a) and (b), we draw tangent lines and measure the slopes of the tangent lines. These values are approximations.

a. For the skydiver:

$$\frac{dv}{dt} \Big|_{(v=0.5)} \approx 1.5$$

$$\frac{dv}{dt} \Big|_{(v=2)} \approx 0.6$$

$$\frac{dv}{dt} \Big|_{(v=1)} \approx 1.2$$

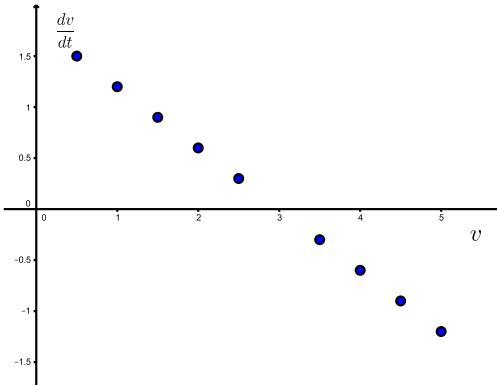
$$\frac{dv}{dt} \Big|_{(v=2.5)} \approx 0.3$$

$$\frac{dv}{dt} \Big|_{(v=1.5)} \approx 0.9$$

b. For the meteorite:

$$\begin{aligned}\left. \frac{dv}{dt} \right|_{(v=3.5)} &\approx -0.3 & \left. \frac{dv}{dt} \right|_{(v=4)} &\approx -0.6 \\ \left. \frac{dv}{dt} \right|_{(v=4.5)} &\approx -0.9 & \left. \frac{dv}{dt} \right|_{(v=5)} &\approx -1.2\end{aligned}$$

A graph of the points from parts (a) and (b) is shown in the following diagram:



- c. Using the values in parts (a) and (b), we obtain $\frac{dv}{dt} = -0.6v + 1.8$.
- d. This differential equation means that the rate of change of velocity with respect to time is a linear function of velocity.
- e. The velocity increases when $\frac{dv}{dt} > 0$. So the velocity increases when $0 < v < 3$.
- f. The velocity decreases when $\frac{dv}{dt} < 0$. So the velocity decreases when $3 < v < 5$.
- g. The velocity will remain constant when $\frac{dv}{dt} = 0$, which is when $v = 3$.

The point of this example is to demonstrate how differential equations model processes in the real world. In this example, two factors influence the velocities: gravity and wind resistance. The differential equation describes how these factors influence the rate of change of the velocities.

7.1.3 Types of Differential Equations

To close this section, we note that differential equations may be classified based on certain characteristics they may possess. You may see many different types of differential equations in a later course in differential equations. For now, we would like to introduce a few terms that are used to describe differential equations.

A *first-order* differential equation is one in which only the first derivative of the function occurs. For this reason,

$$\frac{dv}{dt} = 1.5 - 0.5v$$

is a first-order equation while

$$\frac{d^2y}{dt^2} = -10y$$

is a second-order equation.

A differential equation is *autonomous* if the independent variable does not appear in the description of the derivative. For instance,

$$\frac{dv}{dt} = 1.5 - 0.5v$$

is autonomous because the description of the derivative dv/dt does not depend on time. The equation

$$\frac{dy}{dt} = 1.5t - 0.5y,$$

however, is not autonomous.

7.1.4 Summary

- A differential equation is simply an equation that describes the derivative(s) of an unknown function.
- Physical principles, as well as some everyday situations, often describe how a quantity changes, which lead to differential equations.
- A solution to a differential equation is a function whose derivatives satisfy the equation's description. Differential equations typically have infinitely many solutions, parametrized by the initial values.

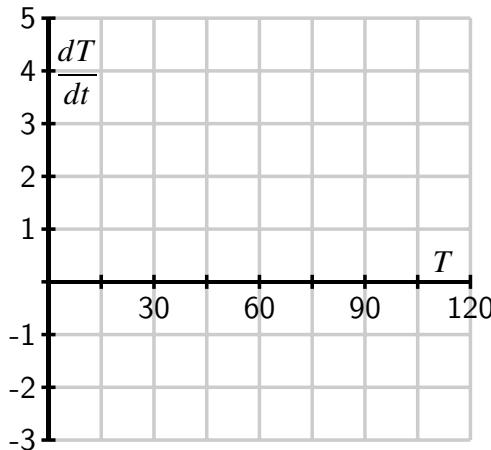
7.1.5 Exercises

1. (Matching solutions with equations) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Finding constant to complete solution) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Choosing solution of $dy/dt = k(1 - Ay)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. Suppose that $T(t)$ represents the temperature of a cup of coffee set out in a room, where T is expressed in degrees Fahrenheit and t in minutes. A physical principle known as Newton's Law of Cooling tells us that

$$\frac{dT}{dt} = -\frac{1}{15}T + 5.$$

- a. Supposes that $T(0) = 105$. What does the differential equation give us for the value of $\frac{dT}{dt}|_{T=105}$? Explain in a complete sentence the meaning of these two facts.

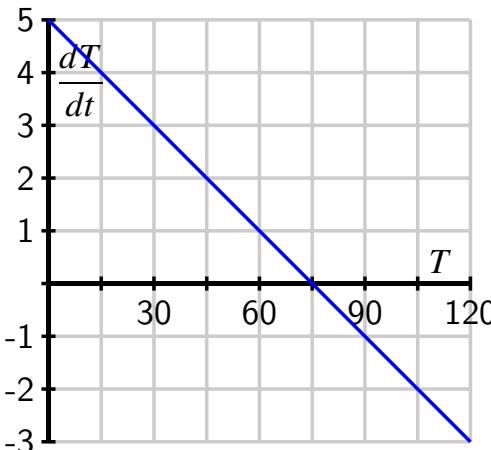
- b. Is T increasing or decreasing at $t = 0$?
- c. What is the approximate temperature at $t = 1$?
- d. On the graph below, make a plot of dT/dt as a function of T .



- e. For which values of T does T increase? For which values of T does T decrease?
- f. What do you think is the temperature of the room? Explain your thinking.
- g. Verify that $T(t) = 75 + 30e^{-t/15}$ is the solution to the differential equation with initial value $T(0) = 105$. What happens to this solution after a long time?

Answer.

- a. $\frac{dT}{dt}|_{T=105} = -2$; when $T = 105$, the coffee's temperature is decreasing at an instantaneous rate of -2 degrees F per minute.
- b. T decreasing at $t = 0$.
- c. $T(1) \approx 103$ degrees F.
- d.



- e. For $T < 75$, T increases. For $T > 75$, T decreases.
- f. Room temperature is 75 degrees F.
- g. Substitute $T(t) = 75 + 30e^{-t/15}$ in for T in the differential equation $\frac{dT}{dt} = -\frac{1}{15}T + 5$ and verify the equality holds; $T(0) = 75 + 30e^0 = 75 + 30 = 105$; $T(t) = 75 + 30e^{-t/15} \rightarrow 75$ as $t \rightarrow \infty$.

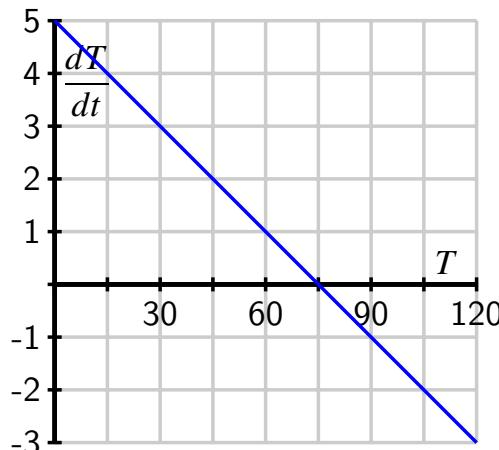
Solution.

- a. Since $\frac{dT}{dt} = -\frac{1}{15}T + 5$ and $T(0) = 105$, $\frac{dT}{dt}|_{T=105} = -\frac{1}{15}(105) + 5 = -7 + 5 = -2$. The fact that $T(0) = 105$ tells us the coffee's temperature at time $t = 0$ is $T = 105$ degrees F, and $\frac{dT}{dt}|_{T=105} = -2$ tells us that at the instant $T = 105$ (i.e. when $t = 0$), the coffee's temperature is decreasing at an instantaneous rate of -2 degrees F per minute.
- b. At $t = 0$, $T = 105$, and since $\frac{dT}{dt}|_{T=105}$ is negative, T decreasing at $t = 0$.
- c. Using the linear approximation $T(1) \approx T(0) + \frac{dT}{dt}|_{t=0}(1 - 0)$ along with the given information, it follows that

$$T(1) \approx 105 + (-2)(1) = 103$$

degrees F.

- d. We note that $\frac{dT}{dt}$ is a linear function of T with slope $-\frac{1}{15}$ and vertical intercept $(0, 5)$, so its graph is as shown below. Note particularly that $\frac{dT}{dt}|_{T=75} = 0$.



- e. For $T < 75$, $\frac{dT}{dt} > 0$, so for these T -values, T increases. For $T > 75$, $\frac{dT}{dt} < 0$, so for these T -values, T decreases.
- f. In the previous question, we observed that 75 degrees F is the critical temperature that changes the behavior of T . There are several reasons to think that 75 degrees F is room temperature: if the coffee's temperature is above room temperature (75), its temperature will decrease to room temperature, while if the coffee is below room temperature, its temperature will increase to room temperature. Moreover, if the coffee's temperature is 75 (which would match the room temperature), then its rate of change is 0 and the coffee's temperature doesn't change.

- g. Verify that $T(t) = 75 + 30e^{-t/15}$ is the solution to the differential equation with initial value $T(0) = 105$. What happens to this solution after a long time? Assuming that $T(t) = 75 + 30e^{-t/15}$, we observe that

$$\frac{dT}{dt} = 30e^{-t/15} \left(-\frac{1}{15} \right) = -2e^{-t/15}.$$

In addition,

$$-\frac{1}{15}T + 5 = -\frac{1}{15}(75 + 30e^{-t/15}) + 5 = -5 - 2e^{-t/15} + 5 = -2e^{-t/15}.$$

The two preceding equations show that

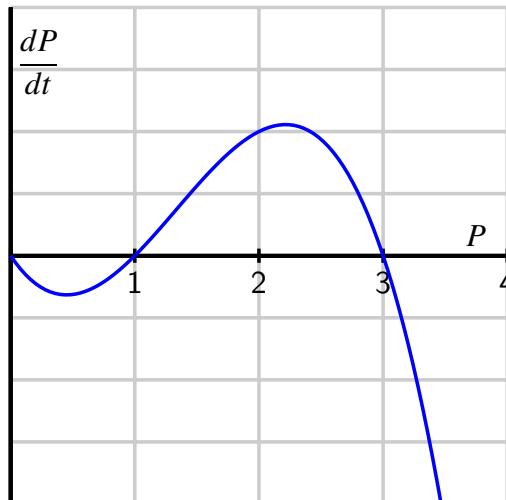
$$\frac{dT}{dt} = -\frac{1}{15}T + 5$$

when we substitute $T(t) = 75 + 30e^{-t/15}$ in for T , and thus this function is a solution to the differential equation. Moreover, $T(0) = 75 + 30e^0 = 75 + 30 = 105$, so the function satisfies the initial condition, too. Finally, because $e^{-t/15} \rightarrow 0$ as $t \rightarrow \infty$, it follows that $T(t) = 75 + 30e^{-t/15} \rightarrow 75$ as $t \rightarrow \infty$, which shows that the solution function's value tends to room temperature as time increases without bound, as we expect.

5. Suppose that the population of a particular species is described by the function $P(t)$, where P is expressed in millions. Suppose further that the population's rate of change is governed by the differential equation

$$\frac{dP}{dt} = f(P)$$

where $f(P)$ is the function graphed below.



- a. For which values of the population P does the population increase?
 b. For which values of the population P does the population decrease?

- c. If $P(0) = 3$, how will the population change in time?
- d. If the initial population satisfies $0 < P(0) < 1$, what will happen to the population after a very long time?
- e. If the initial population satisfies $1 < P(0) < 3$, what will happen to the population after a very long time?
- f. If the initial population satisfies $3 < P(0)$, what will happen to the population after a very long time?
- g. This model for a population's growth is sometimes called "growth with a threshold." Explain why this is an appropriate name.

Answer.

- a. $1 < P < 3$.
- b. $P < 1$ and $3 < P < 4$.
- c. P will not change at all.
- d. The population will decrease toward $P = 0$ with P always being positive.
- e. The population will increase toward $P = 3$ with P always being between 1 and 3.
- f. The population will decrease toward $P = 3$ with P always being above 3.
- g. There's a maximum threshold of $P = 3$.

Solution.

- a. P increases for $1 < P < 3$ since $\frac{dP}{dt} > 0$ for these P -values.
- b. P decreases for $P < 1$ and $3 < P < 4$ since $\frac{dP}{dt} < 0$ for these P -values.
- c. If $P(0) = 3$, $\frac{dP}{dt}|_{P=3} = 0$, so P will not change at all with time since its derivative is 0.
- d. If the initial population satisfies $0 < P(0) < 1$, we expect that the population will decrease toward $P = 0$ with P always being positive. Since $\frac{dP}{dt}$ is always negative for $0 < P < 1$, P is always decreasing on this interval. The population can't drop below $P = 0$ because at that value of P , $\frac{dP}{dt} = 0$ and at that value the population wouldn't change.
- e. If the initial population satisfies $1 < P(0) < 3$, we expect that the population will increase toward $P = 3$ with P always being between 1 and 3. Since $\frac{dP}{dt}$ is always positive for $1 < P < 3$, P is always increasing on this interval. The population can't increase above $P = 3$ because at that value of P , $\frac{dP}{dt} = 0$ and at that value the population wouldn't change.
- f. If the initial population satisfies $3 < P(0)$, we expect that the population will decrease toward $P = 3$ with P always being above 3. Since $\frac{dP}{dt}$ is always negative for $P > 3$, P is always decreasing on this interval. The population can't drop below $P = 3$ because at that value of P , $\frac{dP}{dt} = 0$ and at that value the population wouldn't change.
- g. It looks like there are two thresholds for the population: there's a minimum threshold of $P = 1$, since below that population, the population decreases to 0, and there's a maximum threshold of $P = 3$, where if we are above the minimum but below the maximum, the population works to increase toward the maximum threshold, while if the population is above the maximum, its value will fall back toward it.

6. In this problem, we test further what it means for a function to be a solution to a given differential equation.

- a. Consider the differential equation

$$\frac{dy}{dt} = y - t.$$

Determine whether the following functions are solutions to the given differential equation.

- i. $y(t) = t + 1 + 2e^t$
 - ii. $y(t) = t + 1$
 - iii. $y(t) = t + 2$
- b. When you weigh bananas in a scale at the grocery store, the height h of the bananas is described by the differential equation

$$\frac{d^2h}{dt^2} = -kh$$

where k is the *spring constant*, a constant that depends on the properties of the spring in the scale. After you put the bananas in the scale, you (cleverly) observe that the height of the bananas is given by $h(t) = 4 \sin(3t)$. What is the value of the spring constant?

Answer.

- a. i. $y(t) = t + 1 + 2e^t$ is a solution to the DE.
 ii. $y(t) = t + 1$ is a solution to the DE.
 iii. $y(t) = t + 2$ is a not solution to the DE.
- b. $k = 9$.

Solution.

- a. i. To see if $y(t) = t + 1 + 2e^t$ is a solution to $\frac{dy}{dt} = y - t$, we compute $\frac{dy}{dt}$ and $y - t$ separately and see if they are equal. First, using $y(t) = t + 1 + 2e^t$,

$$\frac{dy}{dt} = 1 + 2e^t.$$

Next,

$$y - t = (t + 1 + 2e^t) - t = 1 + 2e^t.$$

Thus, $y(t) = t + 1 + 2e^t$ satisfies the DE and is indeed a solution.

- ii. For $y(t) = t + 1$, observe that

$$\frac{dy}{dt} = 1$$

and

$$y - t = 1$$

as well, so $y(t) = t + 1$ is a solution to the DE.

- iii. If $y(t) = t + 2$, then

$$\frac{dy}{dt} = 1$$

but

$$y - t = 2$$

as well, so $\frac{dy}{dt} \neq y - t$, and therefore $y(t) = t + 2$ is a not solution to the DE.

- b. In order for $h(t) = 4 \sin(3t)$ to be a solution to

$$\frac{d^2h}{dt^2} = -kh,$$

it must make the equation true. To find k , we substitute the formula for h into the DE and see what is required of k . We observe that for $h(t) = 4 \sin(3t)$,

$$\frac{dh}{dt} = 12 \cos(3t)$$

and

$$\frac{d^2h}{dt^2} = -36 \sin(3t).$$

To satisfy the DE

$$\frac{d^2h}{dt^2} = -kh,$$

it follows that we must have $-36 \sin(3t) = -kh = -k \cdot 4 \sin(3t)$. Thus, we need $-36 = -4k$, and hence $k = 9$.

7.2 Qualitative behavior of solutions to DEs

Motivating Questions

- What is a slope field?
 - How can we use a slope field to obtain qualitative information about the solutions of a differential equation?
 - What are stable and unstable equilibrium solutions of an autonomous differential equation?
-

In earlier work, we have used the tangent line to the graph of a function f at a point a to approximate the values of f near a . The usefulness of this approximation is that we need to know very little about the function; armed with only the value $f(a)$ and the derivative $f'(a)$, we may find the equation of the tangent line and the approximation

$$f(x) \approx f(a) + f'(a)(x - a).$$

Remember that a first-order differential equation gives us information about the derivative of an unknown function. Since the derivative at a point tells us the slope of the tangent line at this point, a differential equation gives us crucial information about the tangent lines to the graph of a solution. We will use this information about the tangent lines to create a *slope field* for the differential equation, which enables us to sketch solutions to initial value problems. Our aim will be to understand the solutions qualitatively. That is, we would like to understand the basic nature of solutions, such as their long-range behavior, without precisely determining the value of a solution at a particular point.

Example 7.2.1. Let's consider the initial value problem

$$\frac{dy}{dt} = t - 2, \quad y(0) = 1.$$

- Use the differential equation to find the slope of the tangent line to the solution $y(t)$ at $t = 0$. Then use the initial value to find the equation of the tangent line at $t = 0$. Sketch this tangent line over the interval $-0.25 \leq t \leq 0.25$ on the axes provided in Figure 7.2.2.

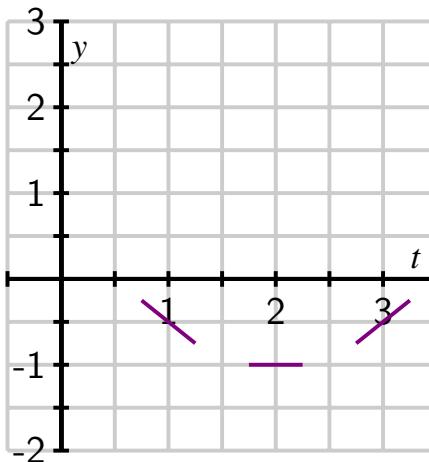


Figure 7.2.2: Grid for plotting partial tangent lines.

- b. Also shown in Figure 7.2.2 are the tangent lines to the solution $y(t)$ at the points $t = 1, 2$, and 3 (we will see how to find these later). Use the graph to measure the slope of each tangent line and verify that each agrees with the value specified by the differential equation.
- c. Using these tangent lines as a guide, sketch a graph of the solution $y(t)$ over the interval $0 \leq t \leq 3$ so that the lines are tangent to the graph of $y(t)$.
- d. Use the Fundamental Theorem of Calculus to find $y(t)$, the solution to this initial value problem.
- e. Graph the solution you found in (d) on the axes provided, and compare it to the sketch you made using the tangent lines.

7.2.1 Slope fields

Example 7.2.1 shows that we can sketch the solution to an initial value problem if we know an appropriate collection of tangent lines. We can use the differential equation to find the slope of the tangent line at any point of interest, and hence plot such a collection.

Let's continue looking at the differential equation $\frac{dy}{dt} = t - 2$. If $t = 0$, this equation says that $dy/dt = 0 - 2 = -2$. Note that this value holds regardless of the value of y . We will therefore sketch tangent lines for several values of y and $t = 0$ with a slope of -2 , as shown in Figure 7.2.3.

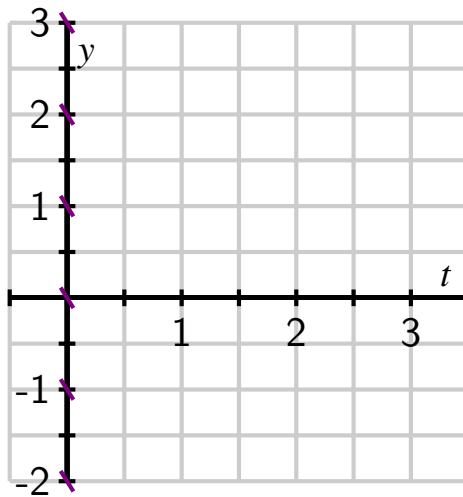


Figure 7.2.3: Tangent lines at points with $t = 0$.

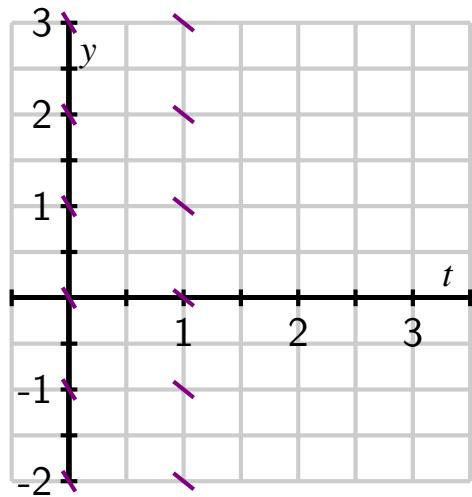


Figure 7.2.4: Adding tangent lines at points with $t = 1$.

Let's continue in the same way: if $t = 2$, the differential equation tells us that $dy/dt = 1 - 2 = -1$, and this holds regardless of the value of y . We now sketch tangent lines for several values of y and $t = 2$ with a slope of -1 in Figure 7.2.4.

Similarly, we see that when $t = 3$, $dy/dt = 0$ and when $t = 4$, $dy/dt = 1$. We may therefore add to our growing collection of tangent line plots to achieve Figure 7.2.5.

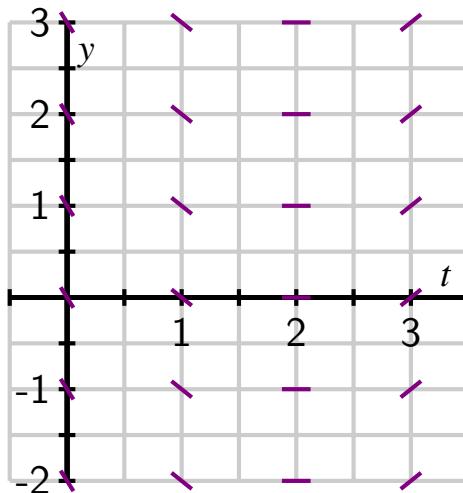


Figure 7.2.5: Adding tangent lines at points with $t = 2$ and $t = 3$.

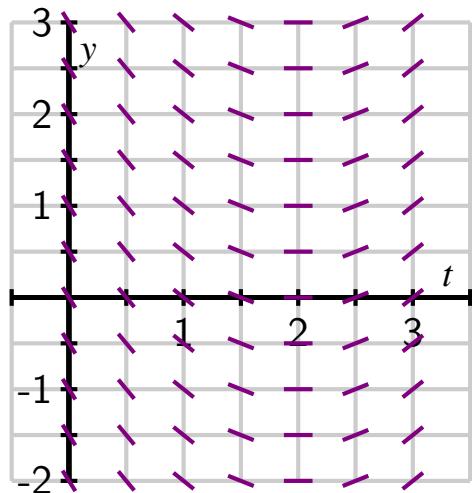


Figure 7.2.6: A completed slope field.

In Figure 7.2.5, we begin to see the solutions to the differential equation emerge. For the sake of even greater clarity, we add more tangent lines to provide the more complete picture

shown at right in Figure 7.2.6.

Figure 7.2.6 is called a *slope field* for the differential equation. It allows us to sketch solutions just as we did in Example 7.2.1. We can begin with the initial value $y(0) = 1$ and start sketching the solution by following the tangent line. Whenever the solution passes through a point at which a tangent line is drawn, that line is tangent to the solution. This principle leads us to the sequence of images in Figure 7.2.7.

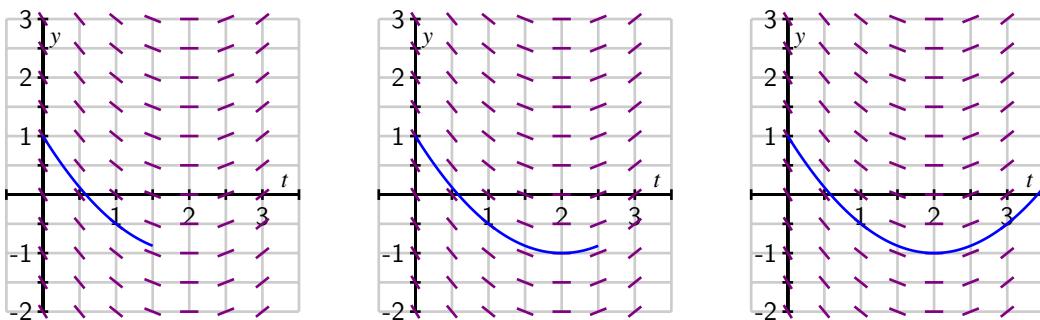


Figure 7.2.7: A sequence of images that show how to sketch the IVP solution that satisfies $y(0) = 1$.

In fact, we can draw solutions for any initial value. Figure 7.2.8 shows solutions for several different initial values for $y(0)$.

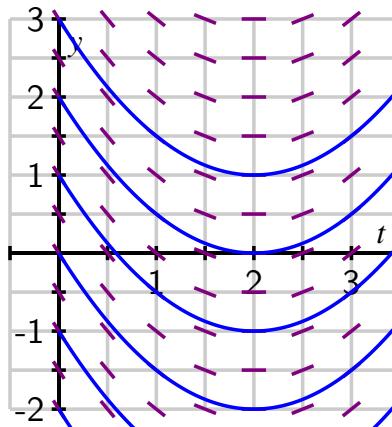


Figure 7.2.8: Different solutions to $\frac{dy}{dt} = t - 2$ that correspond to different initial values.

Just as we did for the equation $\frac{dy}{dt} = t - 2$, we can construct a slope field for any differential equation of interest. The slope field provides us with visual information about how we expect solutions to the differential equation to behave.

Example 7.2.9. Consider the autonomous differential equation

$$\frac{dy}{dt} = -\frac{1}{2}(y - 4).$$

- a. Make a plot of $\frac{dy}{dt}$ versus y on the axes provided in Figure 7.2.10. Looking at the graph, for what values of y does y increase and for what values of y does y decrease?

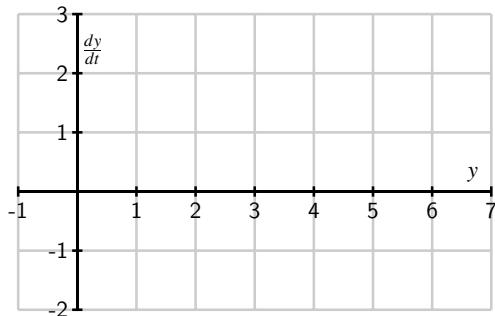


Figure 7.2.10: Axes for plotting $\frac{dy}{dt}$ versus y .

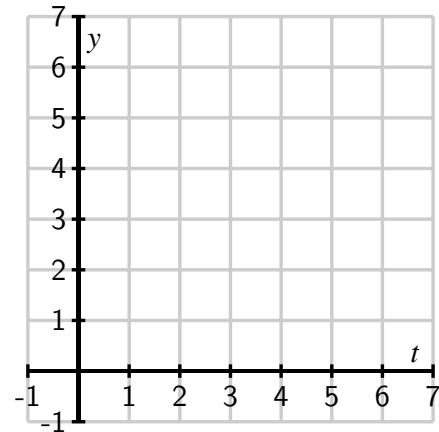
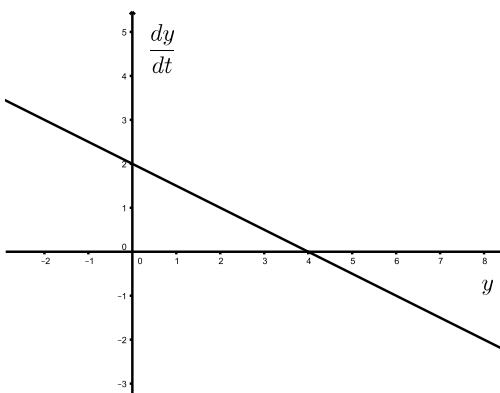


Figure 7.2.11: Axes for plotting the slope field for $\frac{dy}{dt} = -\frac{1}{2}(y - 4)$.

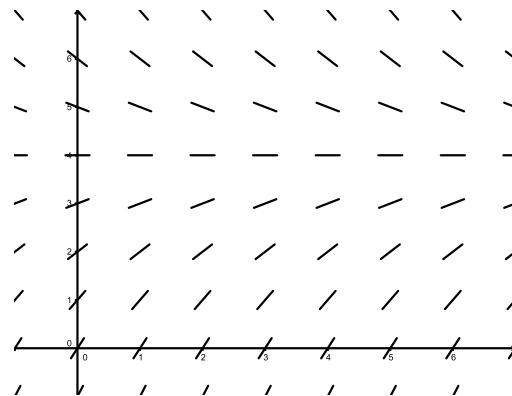
- b. Next, sketch the slope field for this differential equation on the axes provided in Figure 7.2.11.
- c. Use your work in (b) to sketch (on the same axes in Figure 7.2.11.) solutions that satisfy $y(0) = 0$, $y(0) = 2$, $y(0) = 4$ and $y(0) = 6$.
- d. Verify that $y(t) = 4 + 2e^{-t/2}$ is a solution to the given differential equation with the initial value $y(0) = 6$. Compare its graph to the one you sketched in (c).
- e. What is special about the solution where $y(0) = 4$?

Answer.

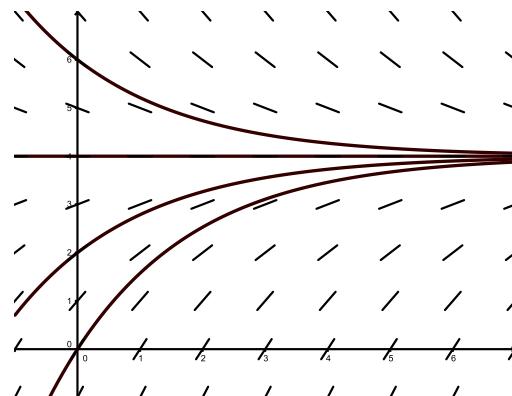
- a. When $y < 4$, y is an increasing function of t . When $y > 4$, y is a decreasing function of t .



b.



c.



d.

$$\frac{dy}{dt} = 2 \left(-\frac{1}{2} e^{-t/2} \right) = -e^{-t/2}$$

and

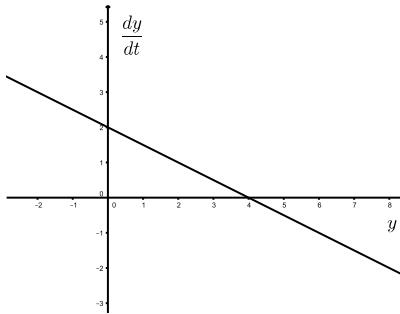
$$-\frac{1}{2}(y - 4) = -\frac{1}{2}(4 + 2e^{-t/2}) = -e^{-t/2}$$

In addition, $y(0) = 4 + 2e^0 = 6$.

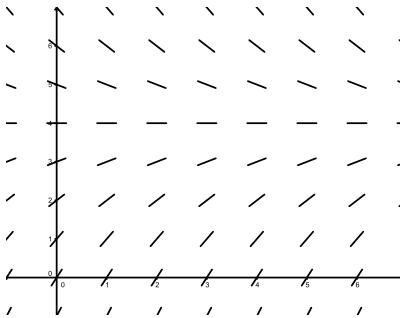
e. A constant function.

Solution.

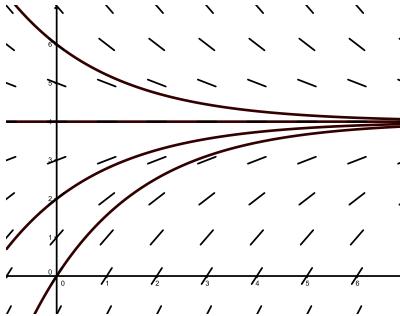
- a. The graph below is a plot of $\frac{dy}{dt}$ versus y . We see that when $y < 4$, $\frac{dy}{dt} > 0$ and hence y is an increasing function of t . We also see that when $y > 4$, $\frac{dy}{dt} < 0$ and hence y is a decreasing function of t .



- b. Next, by sampling points (t, y) in the plane and plotting short lines whose slope is $\frac{dy}{dt}|_{(t,y)}$, we create the following slope field for the differential equation $\frac{dy}{dt} = \frac{1}{2}(y - 4)$.



- c. Finally, by choosing starting points and following the direction of the slope field, we generate plots of solutions that satisfy $y(0) = 0$, $y(0) = 2$, $y(0) = 4$, and $y(0) = 6$, as shown in the following figure.



- d. For $y(t) = 4 + 2e^{-t/2}$, we see that

$$\frac{dy}{dt} = 2 \left(-\frac{1}{2} e^{-t/2} \right) = -e^{-t/2}$$

and

$$-\frac{1}{2}(y - 4) = -\frac{1}{2}(4 + 2e^{-t/2}) = -e^{-t/2}$$

In addition, $y(0) = 4 + 2e^0 = 6$. So $y(t) = 4 + 2e^{-t/2}$ is a solution of the differential equation $\frac{dy}{dt} = \frac{1}{2}(y - 4)$ with $y(0) = 6$. The graph of this function is identical to the top graph in part (c).

- e. The solution where $y(0) = 4$ is a constant function.

7.2.2 Equilibrium solutions and stability

As our work in Example 7.2.9 demonstrates, first-order autonomous equations may have solutions that are constant. These are simple to detect by inspecting the differential equation $dy/dt = f(y)$: constant solutions necessarily have a zero derivative, so $dy/dt = 0 = f(y)$.

For example, in Example 7.2.9, we considered the equation $\frac{dy}{dt} = f(y) = -\frac{1}{2}(y - 4)$. Constant solutions are found by setting $f(y) = -\frac{1}{2}(y - 4) = 0$, which we immediately see implies that $y = 4$.

Values of y for which $f(y) = 0$ in an autonomous differential equation $\frac{dy}{dt} = f(y)$ are called *equilibrium solutions* of the differential equation.

Example 7.2.12. Consider the autonomous differential equation

$$\frac{dy}{dt} = -\frac{1}{2}y(y - 4).$$

- a. Make a plot of $\frac{dy}{dt}$ versus y on the axes provided in Figure 7.2.13. Looking at the graph, for what values of y does y increase and for what values of y does y decrease?

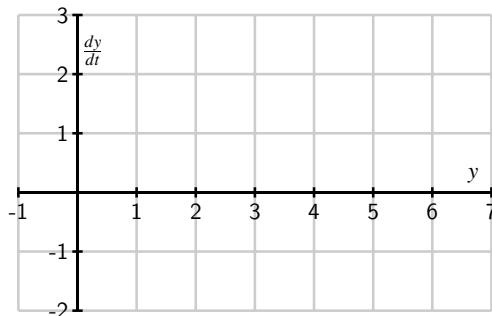


Figure 7.2.13: Axes for plotting dy/dt vs y for $\frac{dy}{dt} = -\frac{1}{2}y(y - 4)$.

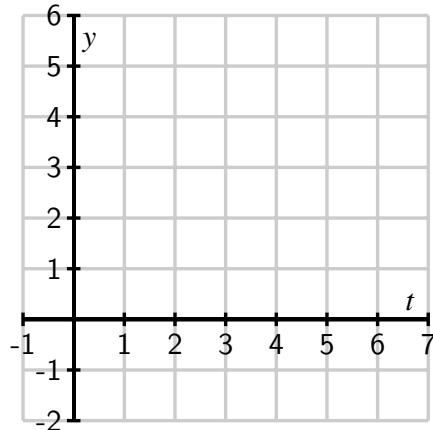


Figure 7.2.14: Axes for plotting the slope field for $\frac{dy}{dt} = -\frac{1}{2}y(y - 4)$.

- b. Identify any equilibrium solutions of the given differential equation.
 c. Now sketch the slope field for the given differential equation on the axes provided in Figure 7.2.14.
 d. Sketch the solutions to the given differential equation that correspond to initial values $y(0) = -1, 0, 1, \dots, 5$.

- e. An equilibrium solution \bar{y} is called *stable* if nearby solutions converge to \bar{y} . This means that if the initial condition varies slightly from \bar{y} , then $\lim_{t \rightarrow \infty} y(t) = \bar{y}$. Conversely, an equilibrium solution \bar{y} is called *unstable* if nearby solutions are pushed away from \bar{y} . Using your work above, classify the equilibrium solutions you found in (b) as either stable or unstable.
- f. Suppose that $y(t)$ describes the population of a species of living organisms and that the initial value $y(0)$ is positive. What can you say about the eventual fate of this population?
- g. Now consider a general autonomous differential equation of the form $dy/dt = f(y)$. Remember that an equilibrium solution \bar{y} satisfies $f(\bar{y}) = 0$. If we graph $dy/dt = f(y)$ as a function of y , for which of the differential equations represented in Figure 7.2.15 and Figure 7.2.16 is \bar{y} a stable equilibrium and for which is \bar{y} unstable? Why?

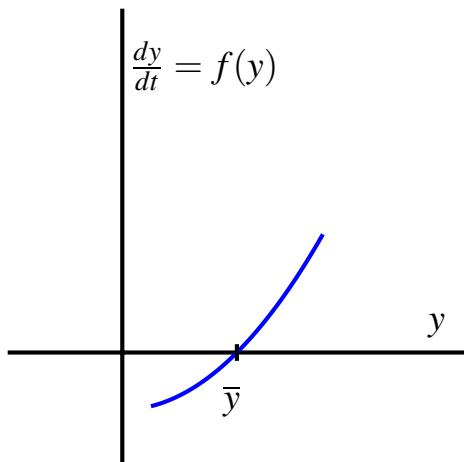


Figure 7.2.15: Plot of $\frac{dy}{dt}$ as a function of y .

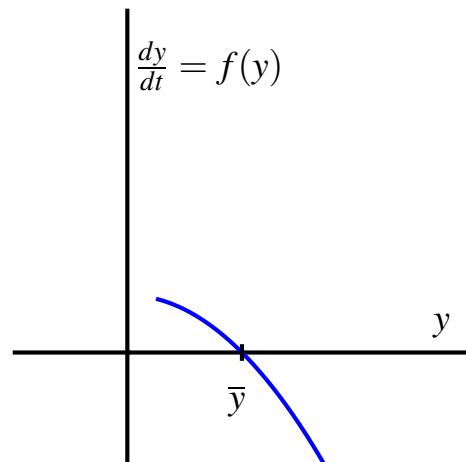
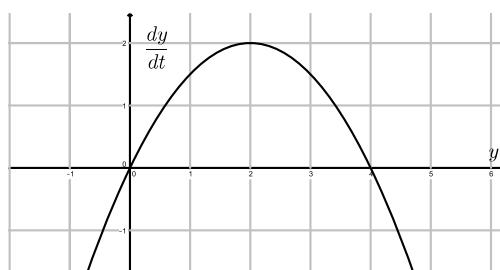


Figure 7.2.16: Plot of $\frac{dy}{dt}$ as a different function of y .

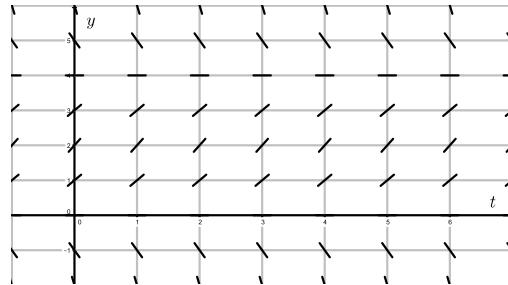
Answer.

- a. When $y < 0$ and when $y > 4$, y is a decreasing function of t . When $0 < y < 4$, y is an increasing function of t .

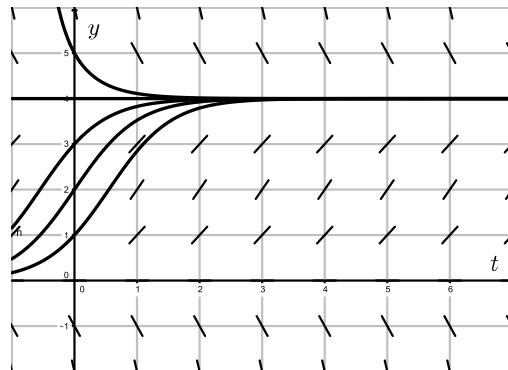


b. $y = 0$ and $y = 4$.

c.



d.



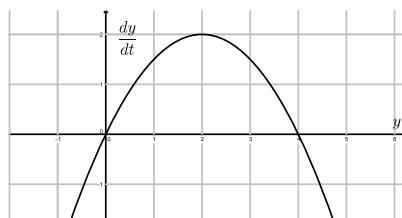
e. $y = 4$ is stable; $y = 0$ is unstable.

f. Tend to 4.

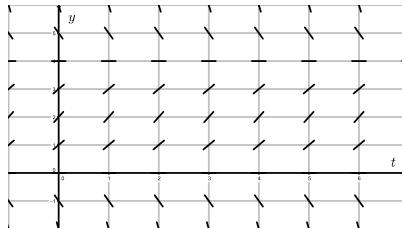
g. Figure 7.2.15 is for an unstable equilibrium; Figure 7.2.16 is for a stable equilibrium.

Solution.

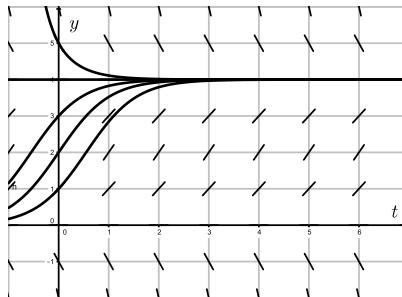
a. The graph below is a plot of $\frac{dy}{dt}$ versus y . We see that when $y < 0$ and when $y > 4$, $\frac{dy}{dt} < 0$ and hence y is a decreasing function of t . We also see that when $0 < y < 4$, $\frac{dy}{dt} > 0$ and hence y is an increasing function of t .



- b. The equilibrium solutions occur when $\frac{dy}{dt} = 0$ and so the equilibrium solutions are $y = 0$ and $y = 4$.
- c. Following is a slope field for the differential equation $\frac{dy}{dt} = -\frac{1}{2}y(y - 4)$.



- d. Below is the slope field for the differential equation $\frac{dy}{dt} = -\frac{1}{2}y(y - 4)$ along with solutions that satisfy $y(0) = 1, y(0) = 2, y(0) = 3, y(0) = 4$, and $y(0) = 5$.



- e. The equilibrium solution $y = 4$ is stable because trajectories nearby approach this constant solution, while the equilibrium solution $y = 0$ is unstable because trajectories nearby veer away from this constant solution.
- f. The population will eventually tend to 4.
- g. Figure 7.2.15 is for an unstable equilibrium. This is because if y is slightly less than \bar{y} , then $\frac{dy}{dt} < 0$ and the function $y(t)$ will be decreasing. In addition, if y is slightly greater than \bar{y} , then $\frac{dy}{dt} > 0$ and the function $y(t)$ will be increasing. In both situations, $y(t)$ is “moving away from” \bar{y} . Figure 7.2.16 is for a stable equilibrium. This is because if y is slightly less than \bar{y} , then $\frac{dy}{dt} > 0$ and the function $y(t)$ will be increasing. In addition, if y is slightly greater than \bar{y} , then $\frac{dy}{dt} < 0$ and the function $y(t)$ will be decreasing. In both situations, $y(t)$ is “moving toward” \bar{y} .

7.2.3 Summary

- A slope field is a plot created by graphing the tangent lines of many different solutions to a differential equation.
- Once we have a slope field, we may sketch the graph of solutions by drawing a curve that is always tangent to the lines in the slope field.

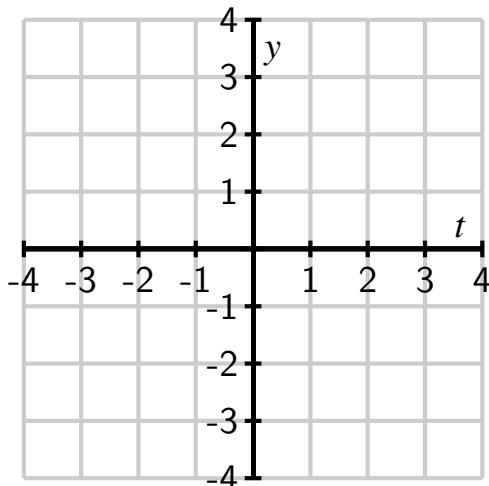
- Autonomous differential equations sometimes have constant solutions that we call equilibrium solutions. These may be classified as stable or unstable, depending on the behavior of nearby solutions.

7.2.4 Exercises

- (Graphing equilibrium solutions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Sketching solution curves) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Matching equations with direction fields) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Describing equilibrium solutions) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- Consider the differential equation

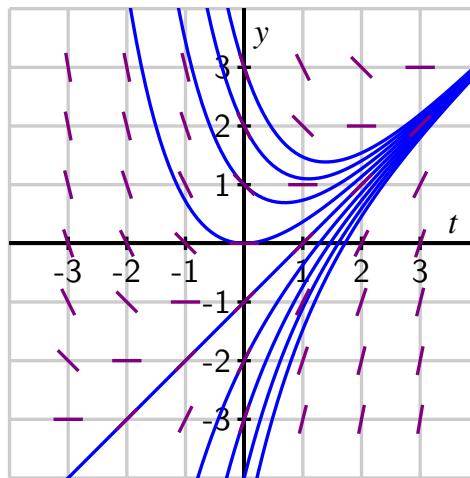
$$\frac{dy}{dt} = t - y.$$

- Sketch a slope field on the axes at right.
- Sketch the solutions whose initial values are $y(0) = -4, -3, \dots, 4$.
- What do your sketches suggest is the solution whose initial value is $y(0) = -1$? Verify that this is indeed the solution to this initial value problem.
- By considering the differential equation and the graphs you have sketched, what is the relationship between t and y at a point where a solution has a local minimum?



Answer.

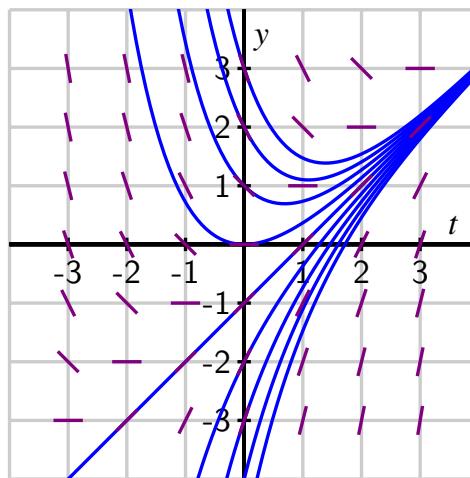
a.



- b. Sketch curves through appropriate points in the slope field above.
- c. $y(t) = t - 1$.
- d. t and y are equal.

Solution.

- a. Sketching the slopes at individual points by computing $\frac{dy}{dt}|_{(t,y)}$ at various values of (t, y) , we see the following slope field.



- b. By following the slope field so that we pass through the points $(0, -4), (0, -3), \dots, (0, 4)$, we see the collection of curves in the figure above.
- c. It looks like the solution that passes through $(0, -1)$ is a straight line with slope $m = 1$, and thus has the formula $y(t) = t - 1$. To verify this function is a solution, we compute both $\frac{dy}{dt}$ and $t - y$. Doing so, we see

$$\frac{dy}{dt} = 1$$

and

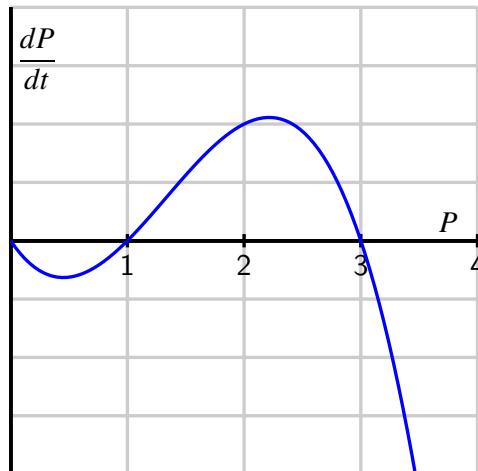
$$t - y = t - (t - 1) = 1.$$

Thus, $y(t) = t - 1$ is indeed a solution to the differential equation.

- d. From both the differential equation and the graphs we have sketched, it appears that t and y are equal at any point where a solution has a local minimum. This is because when $t = y$, $\frac{dy}{dt} = t - y = 0$, and having the derivative be zero is required for any local minimum on a differentiable function.
6. Consider the situation from problem 2 of Section 7.1: Suppose that the population of a particular species is described by the function $P(t)$, where P is expressed in millions. Suppose further that the population's rate of change is governed by the differential equation

$$\frac{dP}{dt} = f(P)$$

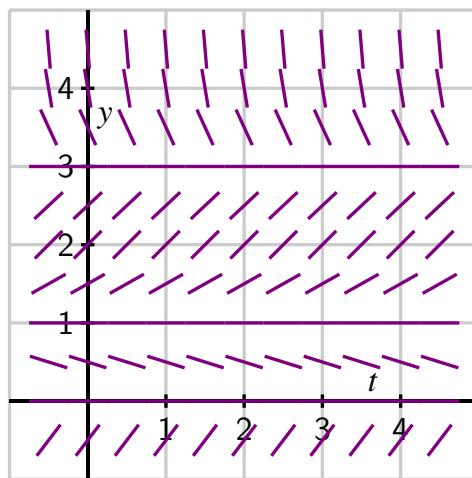
where $f(P)$ is the function graphed below.



- a. Sketch a slope field for this differential equation. You do not have enough information to determine the actual slopes, but you should have enough information to determine where slopes are positive, negative, zero, large, or small, and hence determine the qualitative behavior of solutions.
- b. Sketch some solutions to this differential equation when the initial population $P(0) > 0$.
- c. Identify any equilibrium solutions to the differential equation and classify them as stable or unstable.
- d. If $P(0) > 1$, what is the eventual fate of the species? if $P(0) < 1$?
- e. Remember that we referred to this model for population growth as “growth with a threshold.” Explain why this characterization makes sense by considering solutions whose initial value is close to 1.

Answer.

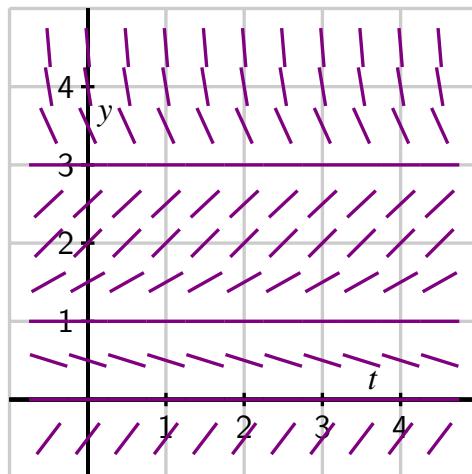
a.



- b. Any solution curve that starts with $P(0) > 3$ will decrease to $P(t) = 3$ as $t \rightarrow \infty$; any curve that starts with $1 < P(0) < 3$ will increase to $P(t) = 3$; any curve that starts with $0 < P(0) < 1$ will decrease to $P(t) = 0$.
- c. $P = 0$, $P = 1$, and $P = 3$. $P = 1$ is unstable; $P = 0$ and $P = 3$ are stable.
- d. The population will stabilize either at the value $P = 3$ or at $P = 0$.
- e. $P(t) = 1$ is the threshold.

Solution.

- a. We first observe from the given plot of $\frac{dP}{dt}$ versus P that $\frac{dP}{dt} = 0$ when $P = 0$, $P = 1$, and $P = 3$. Moreover, we see that $\frac{dP}{dt} > 0$ when $P < 0$ or $1 < P < 3$, and $\frac{dP}{dt} < 0$ otherwise. The closer P is to 0, 1, or 3, the closer $\frac{dP}{dt}$ is to 0. This leads us to sketch a direction field like the following.



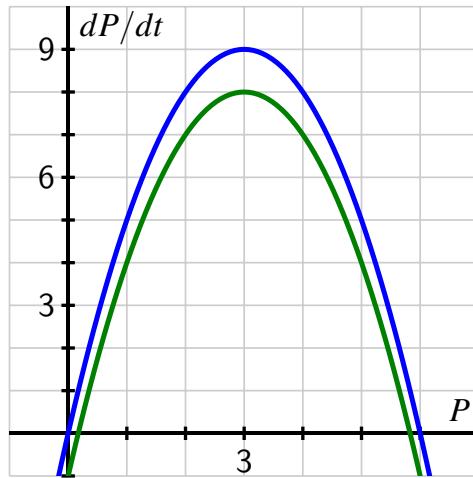
- b. Following the slope field plotted above, we see that any solution curve that starts with $P(0) > 3$ will decrease to $P(t) = 3$ as $t \rightarrow \infty$. Any curve that starts with $1 < P(0) < 3$ will increase to $P(t) = 3$. And any curve that starts with $0 < P(0) < 1$ will decrease to $P(t) = 0$.
- c. The equilibrium solutions are the values of P where $\frac{dP}{dt} = 0$. These are $P = 0$, $P = 1$, and $P = 3$. The equilibrium solution $P = 1$ is unstable, while the solutions $P = 0$ and $P = 3$ are stable.
- d. If $P(0) > 1$, the population's value will tend to $P(t) = 3$ as $t \rightarrow \infty$; if $P(0) < 1$, the population's value will tend to $P(t) = 0$. Thus, the population will stabilize either way, but either at the value $P = 3$ or at a level that corresponds to the population going extinct.
- e. We can view $P(t) = 1$ as the threshold for this population since if there aren't at least $P(t) = 1$ members of the population, the population cannot grow. This aligns well with the definition of the word threshold: "the magnitude or intensity that must be exceeded for a certain reaction, phenomenon, result, or condition to occur or be manifested".
7. The population of a species of fish in a lake is $P(t)$ where P is measured in thousands of fish and t is measured in months. The growth of the population is described by the differential equation

$$\frac{dP}{dt} = f(P) = P(6 - P).$$

- a. Sketch a graph of $f(P) = P(6 - P)$ and use it to determine the equilibrium solutions and whether they are stable or unstable. Write a complete sentence that describes the long-term behavior of the fish population.
- b. Suppose now that the owners of the lake allow fishers to remove 1000 fish from the lake every month (remember that $P(t)$ is measured in *thousands* of fish). Modify the differential equation to take this into account. Sketch the new graph of dP/dt versus P . Determine the new equilibrium solutions and decide whether they are stable or unstable.
- c. Given the situation in part (b), give a description of the long-term behavior of the fish population.
- d. Suppose that fishermen remove h thousand fish per month. How is the differential equation modified?
- e. What is the largest number of fish that can be removed per month without eliminating the fish population? If fish are removed at this maximum rate, what is the eventual population of fish?

Answer.

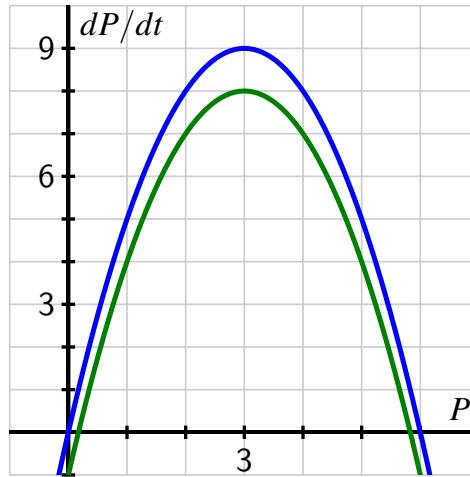
- a. A graph of f against P is given in blue in the figure below. The equilibrium solutions are $P = 0$ (unstable) and $P = 6$ (stable).



- b. $\frac{dP}{dt} = g(P) = P(6 - P) - 1$; the equilibrium at $P \approx 0.172$ is unstable; the equilibrium at $P \approx 5.83$ is stable.
- c. If $P < \frac{6-\sqrt{32}}{2}$, then the fish population will die out. If $\frac{6-\sqrt{32}}{2} < P < \frac{6+\sqrt{32}}{2}$, then the fish population will approach $\frac{6+\sqrt{32}}{2}$ thousand fish.
- d. $\frac{dP}{dt} = g(P) = P(6 - P) - h$; equilibrium solutions $P = \frac{6+\sqrt{36-4h}}{2}, \frac{6-\sqrt{36-4h}}{2}$.
- e. 9000 fish; harvesting at that rate will maintain the number of fish we start with, provided it's at least 3000.

Solution.

- a. A graph of f against P is given in blue in the figure below. This graph and the differential equation show that the equilibrium solutions occur when $f(P) = 0$, or when $P = 0$ or $P = 6$. Since $P' < 0$ for P slightly less than 0, the solutions with $P < 0$ decrease away from the equilibrium $P = 0$. Also, $P' > 0$ for P slightly larger than 0 means that the solutions for $P > 0$ increase away from the equilibrium $P = 0$. So $P = 0$ is an unstable equilibrium. If P is slightly less than 6 we see that $P > 0$, so these solutions increase toward the equilibrium $P = 0$. When P is slightly larger than 6, we have $P' < 0$ and so these solutions decrease toward the equilibrium $P = 6$. We conclude that the equilibrium solution $P = 6$ is stable.



- b. Since P will decrease by 1 thousand fish every month through harvesting, the new differential equation is

$$\frac{dP}{dt} = g(P) = P(6 - P) - 1.$$

A sketch of g against P is given in green in the figure in (a). The equilibrium solutions now occur when $g(P) = P(6 - P) - 1 = 0$. This happens when $P^2 - 6P + 1 = 0$. The quadratic formula tells us that the solutions to this quadratic equation are

$$P = \frac{6 + \sqrt{32}}{2} \approx 5.83 \quad \text{and} \quad P = \frac{6 - \sqrt{32}}{2} \approx 0.172.$$

The same argument as in (a) shows that the equilibrium at $P \approx 0.172$ is unstable and the equilibrium at $P \approx 5.83$ is stable.

- c. Assuming that there are fish in the lake, so that $P > 0$, there are two cases to consider. If $P < \frac{6 - \sqrt{32}}{2}$, then $P' < 0$ and the fish population will die out. If $\frac{6 - \sqrt{32}}{2} < P < \frac{6 + \sqrt{32}}{2}$, then $P' > 0$ and the fish population will increase and approach a stable equilibrium population of $\frac{6 + \sqrt{32}}{2}$ thousand fish. If $P > \frac{6 + \sqrt{32}}{2}$, then $P' < 0$ and the fish population will decrease and approach a stable equilibrium population of $\frac{6 + \sqrt{32}}{2}$ thousand fish.
- d. In this case the differential equation becomes

$$\frac{dP}{dt} = g(P) = P(6 - P) - h.$$

We will have equilibrium solutions when $P^2 - 6P + h = 0$ or when

$$P = \frac{6 + \sqrt{36 - 4h}}{2} \quad \text{and} \quad P = \frac{6 - \sqrt{36 - 4h}}{2}.$$

- e. We need a positive stable equilibrium. This will happen as long as $36 - 4h \geq 0$ or when $h \leq 9$. So the maximum harvest is 9000 fish. Note that P' is negative, so we need to start with more than 3000 fish. Since $P(6 - P)$ has a maximum value of 9, the fish will reproduce quickly enough so that harvesting 9000 fish will maintain a stable population.

8. Let $y(t)$ be the number of thousands of mice that live on a farm; assume time t is measured in years.¹

- The population of the mice grows at a yearly rate that is twenty times the number of mice. Express this as a differential equation.
- At some point, the farmer brings C cats to the farm. The number of mice that the cats can eat in a year is

$$M(y) = C \frac{y}{2+y}$$

thousand mice per year. Explain how this modifies the differential equation that you found in part a).

- Sketch a graph of the function $M(y)$ for a single cat $C = 1$ and explain its features by looking, for instance, at the behavior of $M(y)$ when y is small and when y is large.
- Suppose that $C = 1$. Find the equilibrium solutions and determine whether they are stable or unstable. Use this to explain the long-term behavior of the mice population depending on the initial population of the mice.
- Suppose that $C = 60$. Find the equilibrium solutions and determine whether they are stable or unstable. Use this to explain the long-term behavior of the mice population depending on the initial population of the mice.
- What is the smallest number of cats you would need to keep the mice population from growing arbitrarily large?

Answer.

- $\frac{dy}{dt} = 20y$.
- $\frac{dy}{dt} = 20y - C \frac{y}{2+y}$
- For positive y near 0, $M(y) = \frac{y}{2+y} \approx 0$; for large values of y , $M(y) = \frac{y}{2+y} \approx 1$.
- The only equilibrium solution is $y = 0$, which is unstable.
- The equilibrium solutions are $y = 0$ (stable) and $y = 1$ (unstable).
- At least 41 cats.

Solution.

- If we let $y(t)$ be the population of the mice at year t in thousands and it grows at a yearly rate that is twenty times the number of mice, the governing differential equation is $\frac{dy}{dt} = 20y$.
- On their own, the mice add $20y$ thousand mice per year to their total population. With C cats present, the cats now remove $M(y) = C \frac{y}{2+y}$ thousand mice per year. By taking both the rates at which mice are added and removed from the population into account, we find the updated differential equation

$$\frac{dy}{dt} = 20y - C \frac{y}{2+y}$$

¹This problem is based on an ecological analysis presented in a research paper by C.S. Hollings: The Components of Predation as Revealed by a Study of Small Mammal Predation of the European Pine Sawfly, *Canadian Entomology* 91: 283-320.

- c. For positive y near 0, $M(y) = \frac{y}{2+y} \approx 0$; for large values of y , $M(y) = \frac{y}{2+y} \approx 1$. This means that a single cat can remove about 1 thousand mice per year from the population when the mice population is large.
- d. Letting $C = 1$, we are working with the differential equation

$$\frac{dy}{dt} = 20y - \frac{y}{2+y}.$$

To find the equilibrium solutions, we first set $\frac{dy}{dt} = 0$. Doing so, we find that $20y = \frac{y}{2+y}$, so $40y + 20y^2 = y$, or $39y + 20y^2 = 0$. Factoring, we see $y(39y + 20) = 0$ and therefore $y = 0$ or $y = -\frac{20}{39}$. Since a negative equilibrium solution doesn't make sense in the context of the problem, the only equilibrium solution to study is $y = 0$. If we plot dy/dt versus y for $y > 0$, we see that dy/dt is always positive, and thus the population will grow away from $y = 0$, making this equilibrium unstable. And this makes sense: the population of mice will grow without bound, because one cat is insufficient to keep up with a mice population growing at a rate of 20 times its size per year.

- e. Letting $C = 60$, we are working with the differential equation

$$\frac{dy}{dt} = 20y - \frac{60y}{2+y}.$$

Similar work as in (d) shows that the equilibrium solutions are $y = 0$ or $y = 1$, and if we plot dy/dt versus y , we see that for $0 < y < 1$, $dy/dt < 0$, while for $y > 1$, $dy/dt > 0$. Thus, with 60 cats present, as long as there are initially fewer than 1000 mice, the cats will drive the mice population to 0. If there are more than 1000 mice initially present, the mice population will grow without bound. And if there are 1000 mice initially present, this will be an (unstable) equilibrium.

- f. In order to keep the mice population from growing arbitrarily large, we need to bring enough cats that for certain initial mice populations, we can make $dy/dt \leq 0$. In other words, we need enough cats that there's a positive equilibrium solution. Considering the general differential equation,

$$\frac{dy}{dt} = 20y - C \frac{y}{2+y},$$

we thus set $dy/dt = 0$ and solve for C . Doing so, we find that $0 = 20y - C \frac{y}{2+y}$, so $20y(y+2) = Cy$ or $20y^2 + 40y - Cy = 0$. Factoring, we find that

$$y(20y + (40 - C)) = 0.$$

To have a positive equilibrium solution, we require that $40 - C < 0$, and therefore the farmer must bring at least 41 cats to the farm. Even in that circumstance, whether the cats can keep the number of mice from growing without bound will depend on the initial number of mice present.

7.3 Euler's method

Motivating Questions

- What is Euler's method and how can we use it to approximate the solution to an initial value problem?
 - How accurate is Euler's method?
-

In Section 7.2, we saw how a slope field can be used to sketch solutions to a differential equation. In particular, the slope field is a plot of a large collection of tangent lines to a large number of solutions of the differential equation, and we sketch a single solution by simply following these tangent lines. With a little more thought, we can use this same idea to approximate numerically the solutions of a differential equation.

Example 7.3.1. Consider the initial value problem

$$\frac{dy}{dt} = \frac{1}{2}(y + 1), \quad y(0) = 0.$$

- Use the differential equation to find the slope of the tangent line to the solution $y(t)$ at $t = 0$. Then use the given initial value to find the equation of the tangent line at $t = 0$.
- Sketch the tangent line on the axes provided in Figure 7.3.2 on the interval $0 \leq t \leq 2$ and use it to approximate $y(2)$, the value of the solution at $t = 2$.

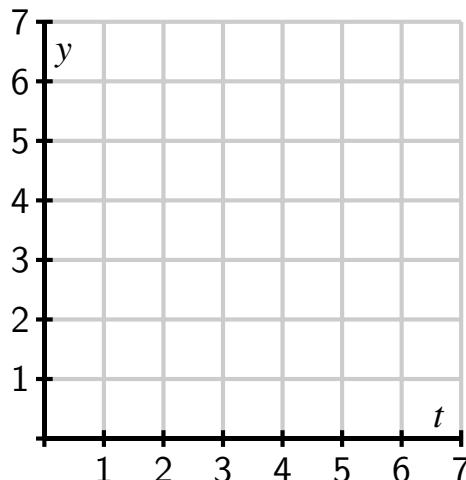


Figure 7.3.2: Grid for plotting the tangent line.

- c. Assuming that your approximation for $y(2)$ is the actual value of $y(2)$, use the differential equation to find the slope of the tangent line to $y(t)$ at $t = 2$. Then, write the equation of the tangent line at $t = 2$.
- d. Add a sketch of this tangent line on the interval $2 \leq t \leq 4$ to your plot Figure 7.3.2; use this new tangent line to approximate $y(4)$, the value of the solution at $t = 4$.
- e. Repeat the same step to find an approximation for $y(6)$.

7.3.1 Euler's Method

Example 7.3.1 demonstrates an algorithm known as Euler's¹ Method, which generates a numerical approximation to the solution of an initial value problem. In this algorithm, we will approximate the solution by taking horizontal steps of a fixed size that we denote by Δt .

[stack]

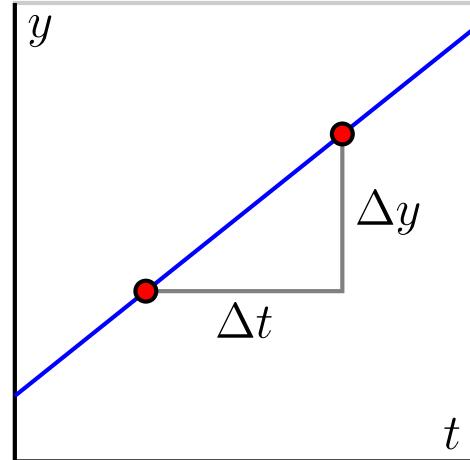


Figure 7.3.3: The role of slope in Euler's Method.

Now, suppose that we would like to solve the initial value problem

$$\frac{dy}{dt} = t - y, \quad y(0) = 1.$$

There is an algorithm by which we can find an algebraic formula for the solution to this initial value problem, and we can check that this solution is $y(t) = t - 1 + 2e^{-t}$. But we are instead interested in generating an approximate solution by creating a sequence of points (t_i, y_i) , where $y_i \approx y(t_i)$. For this first example, we choose $\Delta t = 0.2$.

Since we know that $y(0) = 1$, we will take the initial point to be $(t_0, y_0) = (0, 1)$ and move horizontally by $\Delta t = 0.2$ to the point (t_1, y_1) . Thus, $t_1 = t_0 + \Delta t = 0.2$. Now, the differential

¹"Euler" is pronounced "Oy-ler." Among other things, Euler is the mathematician credited with the famous number e ; if you incorrectly pronounce his name "You-ler," you fail to appreciate his genius and legacy.

equation tells us that the slope of the tangent line at this point is

$$m = \frac{dy}{dt} \Big|_{(0,1)} = 0 - 1 = -1,$$

so to move along the tangent line by taking a horizontal step of size $\Delta t = 0.2$, we must also move vertically by

$$\Delta y = m\Delta t = -1 \cdot 0.2 = -0.2.$$

We then have the approximation $y(0.2) \approx y_1 = y_0 + \Delta y = 1 - 0.2 = 0.8$. At this point, we have executed one step of Euler's method, as seen graphically in Figure 7.3.4.

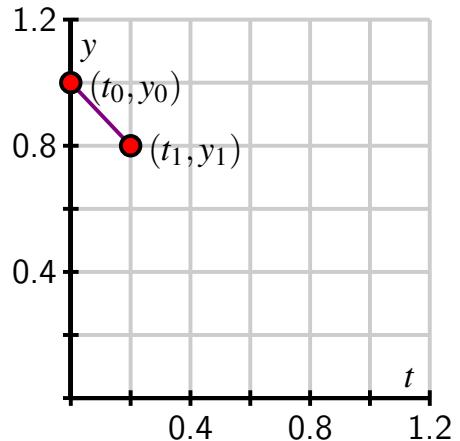


Figure 7.3.4: One step of Euler's method.

Now we repeat this process: at $(t_1, y_1) = (0.2, 0.8)$, the differential equation tells us that the slope is

$$m = \frac{dy}{dt} \Big|_{(0.2, 0.8)} = 0.2 - 0.8 = -0.6.$$

If we move forward horizontally by Δt to $t_2 = t_1 + \Delta t = 0.4$, we must move vertically by

$$\Delta y = -0.6 \cdot 0.2 = -0.12.$$

We consequently arrive at $y_2 = y_1 + \Delta y = 0.8 - 0.12 = 0.68$, which gives $y(0.4) \approx 0.68$. Now we have completed the second step of Euler's method, as shown in Figure 7.3.5.

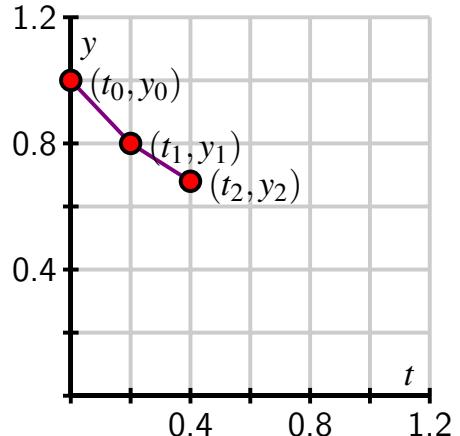


Figure 7.3.5: Two steps of Euler's method.

If we continue in this way, we may generate the points (t_i, y_i) shown in Figure 7.3.6. Be-

cause we can find a formula for the actual solution $y(t)$ to this differential equation, we can graph $y(t)$ and compare it to the points generated by Euler's method, as shown in Figure 7.3.7.

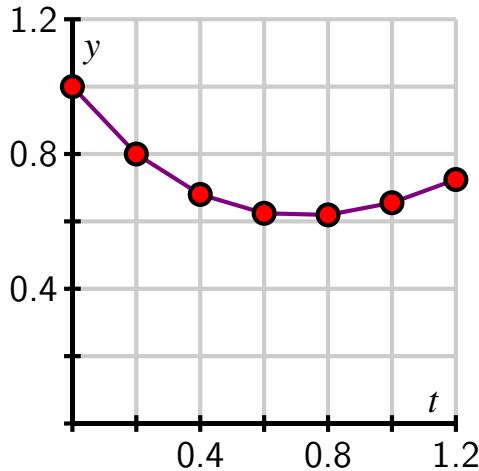


Figure 7.3.6: The points and piecewise linear approximate solution generated by Euler's method.

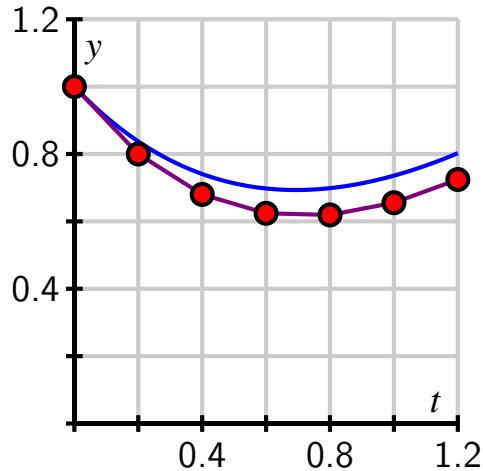


Figure 7.3.7: The approximate solution compared to the exact solution (shown in blue).

Because we need to generate a large number of points (t_i, y_i) , it is convenient to organize the implementation of Euler's method in a table as shown. We begin with the given initial data.

t_i	y_i	dy/dt	Δy
0.0000	1.0000		

From here, we compute the slope of the tangent line $m = dy/dt$ using the formula for dy/dt from the differential equation, and then we find Δy , the change in y , using the rule $\Delta y = m\Delta t$.

t_i	y_i	dy/dt	Δy
0.0000	1.0000	-1.0000	-0.2000

Next, we increase t_i by Δt and y_i by Δy to get

t_i	y_i	dy/dt	Δy
0.0000	1.0000	-1.0000	-0.2000
0.2000	0.8000		

We continue the process for however many steps we decide, eventually generating a table like Table 7.3.8.

t_i	y_i	dy/dt	Δy
0.0000	1.0000	-1.0000	-0.2000
0.2000	0.8000	-0.6000	-0.1200
0.4000	0.6800	-0.2800	-0.0560
0.6000	0.6240	-0.0240	-0.0048
0.8000	0.6192	0.1808	0.0362
1.0000	0.6554	0.3446	0.0689
1.2000	0.7243	0.4757	0.0951

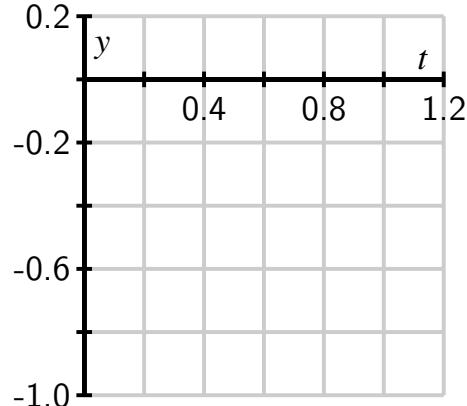
Table 7.3.8: Euler's method for 6 steps with $\Delta t = 0.2$.

Example 7.3.9. Consider the initial value problem

$$\frac{dy}{dt} = 2t - 1, \quad y(0) = 0$$

- a. Use Euler's method with $\Delta t = 0.2$ to approximate the solution at $t_i = 0.2, 0.4, 0.6, 0.8$, and 1.0. Record your work in the following table, and sketch the points (t_i, y_i) on the axes provided.

t_i	y_i	dy/dt	Δy
0.0000	0.0000		
0.2000			
0.4000			
0.6000			
0.8000			
1.0000			

**Table 7.3.10:** Table for recording results of Euler's method.**Figure 7.3.11:** Grid for plotting points generated by Euler's method.

- b. Find the exact solution to the original initial value problem and use this function to find the error in your approximation at each one of the points t_i .
- c. Explain why the value y_5 generated by Euler's method for this initial value problem produces the same value as a left Riemann sum for the definite integral $\int_0^1 (2t - 1) dt$.
- d. How would your computations differ if the initial value was $y(0) = 1$? What does this mean about different solutions to this differential equation?

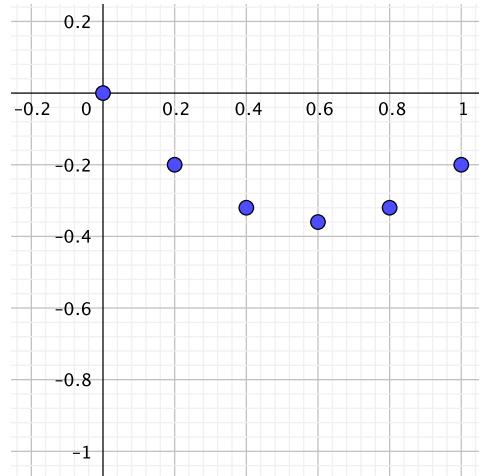
Hint.

- a. Small hints for each of the prompts above.

Answer.

a.

t_i	y_i	dy/dt	Δy
0	0	-1	-0.2
0.2	-0.2	-0.6	-0.12
0.4	-0.32	-0.2	-0.04
0.6	-0.36	0.2	0.04
0.8	-0.32	0.6	0.12
1.0	-0.2	1	0.2



- b. $y = t^2 - t$, with errors $e_1 = 0.04, e_2 = 0.08, e_3 = 0.12, e_4 = 0.16, e_5 = 0.2$.

- c. If we first think about how y_1 is generated for the initial value problem $\frac{dy}{dt} = f(t) = 2t - 1, y(0) = 0$, we see that $y_1 = y_0 + \Delta t \cdot f(t_0)$. Since $y_0 = 0$, we have $y_1 = \Delta t \cdot f(t_0)$. From there, we know that y_2 is given by $y_2 = y_1 + \Delta t f(t_1)$. Substituting our earlier result for y_1 , we see that $y_2 = \Delta t \cdot f(t_0) + \Delta t f(t_1)$. Continuing this process up to y_5 , we get

$$y_5 = \Delta t \cdot f(t_0) + \Delta t f(t_1) + \Delta t f(t_2) + \Delta t f(t_3) + \Delta t f(t_4)$$

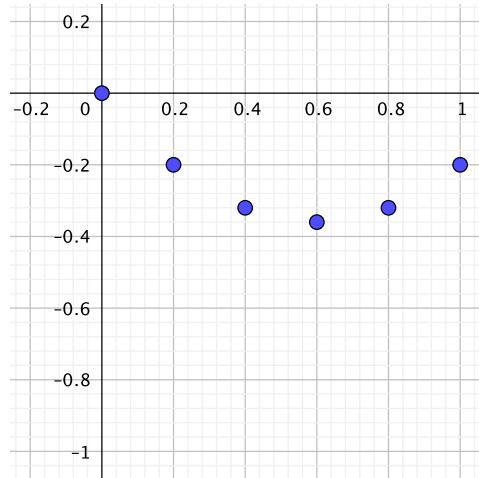
This is precisely the left Riemann sum with five subintervals for the definite integral $\int_0^1 (2t - 1) dt$.

- d. Solutions to this differential equation all differ by only a constant.

Solution.

- a. Using Euler's Method with $\Delta t = 0.2$, we find the results shown in the table and graph below.

t_i	y_i	dy/dt	Δy
0	0	-1	-0.2
0.2	-0.2	-0.6	-0.12
0.4	-0.32	-0.2	-0.04
0.6	-0.36	0.2	0.04
0.8	-0.32	0.6	0.12
1.0	-0.2	1	0.2



- b. The exact solution to the initial value problem $\frac{dy}{dt} = 2t - 1$, $y(0) = 0$ is $y = t^2 - t$, which we find by integrating both sides with respect to t and then applying the fact that $y(0) = 0$ to determine that the integration constant is 0. From here, we can see that the errors are $e_1 = y(0.2) - y_1 = 0.04$, $e_2 = y(0.4) - y_2 = 0.08$, $e_3 = y(0.6) - y_3 = 0.12$, $e_4 = y(0.8) - y_4 = 0.16$, $e_5 = y(1.0) - y_5 = 0.2$.
- c. If we first think about how y_1 is generated for the initial value problem $\frac{dy}{dt} = f(t) = 2t - 1$, $y(0) = 0$, we see that $y_1 = y_0 + \Delta t \cdot f(t_0)$. Since $y_0 = 0$, we have $y_1 = \Delta t \cdot f(t_0)$. From there, we know that y_2 is given by $y_2 = y_1 + \Delta t f(t_1)$. Substituting our earlier result for y_1 , we see that $y_2 = \Delta t \cdot f(t_0) + \Delta t f(t_1)$. Continuing this process up to y_5 , we get

$$y_5 = \Delta t \cdot f(t_0) + \Delta t f(t_1) + \Delta t f(t_2) + \Delta t f(t_3) + \Delta t f(t_4)$$

This is precisely the left Riemann sum with five subintervals for the definite integral $\int_0^1 (2t - 1) dt$.

- d. Since $\frac{dy}{dt} = 2t - 1$, the slope of the tangent line depends only on t . Thus, the different initial condition of $y(0) = 1$ simply adds 1 to every value of y_i in our computations, which shifts every point we compute by Euler's Method up by 1 unit. This tells us that the solutions to this differential equation all differ by only a constant. We can see that both in our computations with Euler's Method and if we solve the differential equation to get $y = t^2 - t + C$.

Example 7.3.12. Consider the differential equation $\frac{dy}{dt} = 6y - y^2$.

- a. Sketch the slope field for this differential equation on the axes provided in Figure 7.3.13.

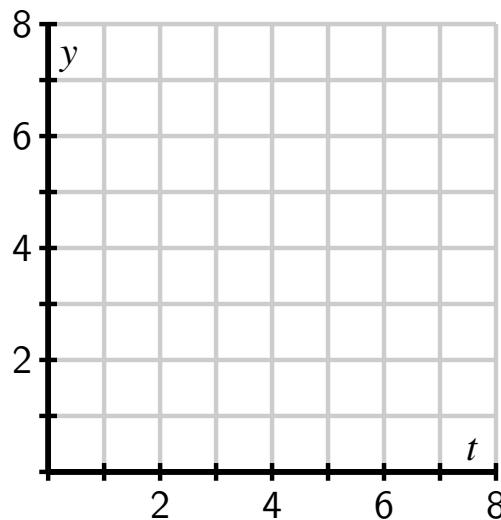


Figure 7.3.13: Grid for plotting the slope field of the given differential equation.

- b. Identify any equilibrium solutions and determine whether they are stable or unstable.
- c. What is the long-term behavior of the solution that satisfies the initial value $y(0) = 1$?
- d. Using the initial value $y(0) = 1$, use Euler's method with $\Delta t = 0.2$ to approximate the solution at $t_i = 0.2, 0.4, 0.6, 0.8$, and 1.0 . Record your results in Table 7.3.14 and sketch the corresponding points (t_i, y_i) on the axes provided in Figure 7.3.15. Note the different horizontal scale on the axes in Figure 7.3.15 compared to Figure 7.3.13.

t_i	y_i	dy/dt	Δy
0.0	1.0000		
0.2			
0.4			
0.6			
0.8			
1.0			

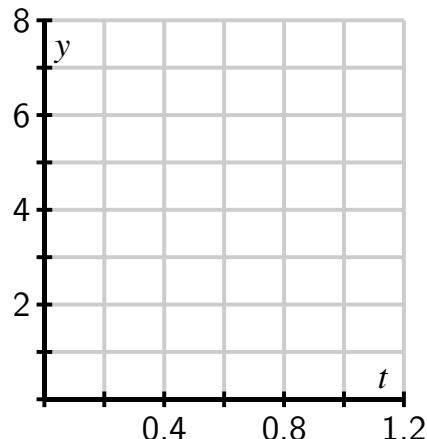


Table 7.3.14: Table for recording results of Euler's method with $\Delta t = 0.2$.

Figure 7.3.15: Axes for plotting the results of Euler's method.

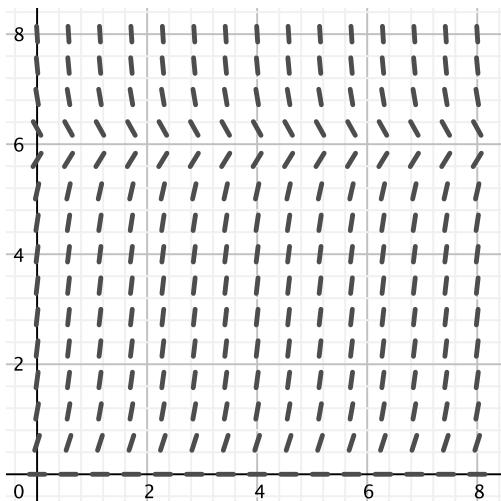
- e. What happens if we apply Euler's method to approximate the solution with $y(0) = 6$?

Hint.

- a. Small hints for each of the prompts above.

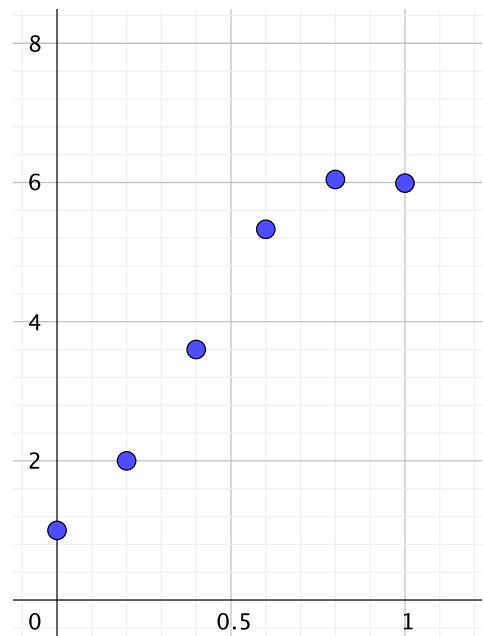
Answer.

a.



- b. $y = 0$ or $y = 6$.
 c. The solution will tend to $y = 6$.
 d.

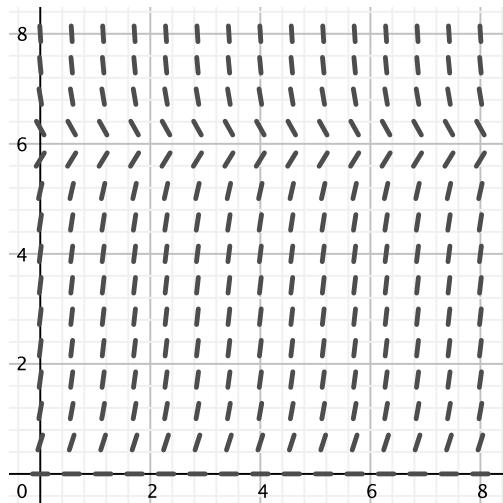
t_i	y_i	dy/dt	Δy
0.0	1.0000	5.0000	1.0000
0.2	2.0000	8.0000	1.6000
0.4	3.6000	8.6400	1.7280
0.6	5.3280	3.5804	0.7161
0.8	6.0441	-0.2664	-0.0533
1.0	5.9908	0.0551	0.0110



- e. The value of $y_i = 6$ for every value of i .

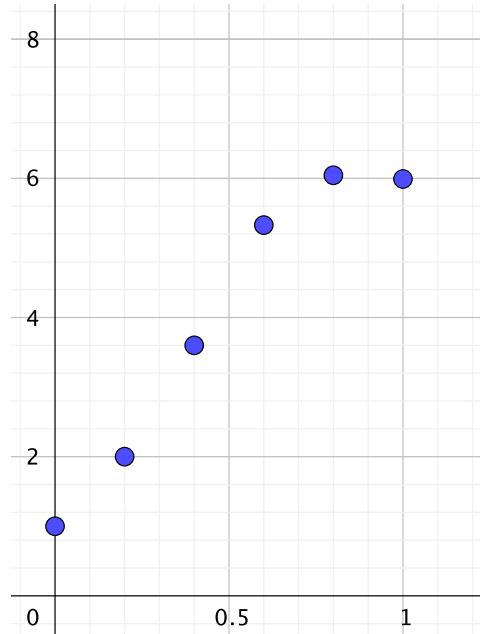
Solution.

- a. The slope field for this differential equation is shown in the following figure.



- b. If y is an equilibrium solution, then y is constant so $\frac{dy}{dt} = 0$. Solving $0 = 6y - y^2$, we see that either $y = 0$ or $y = 6$, which is also consistent with what we see in the slope field.
- c. The long-term behavior of the solution that satisfies the initial value $y(0) = 1$ is that the solution will tend to $y = 6$.
- d. Using Euler's method with $\Delta t = 0.2$, we approximate the solution of the IVP at $t_i = 0.2, 0.4, 0.6, 0.8$, and 1.0 as shown in the following table and figure. We see that our estimates actually jump slightly above the equilibrium solution before being pulled back down.

t_i	y_i	dy/dt	Δy
0.0	1.0000	5.0000	1.0000
0.2	2.0000	8.0000	1.6000
0.4	3.6000	8.6400	1.7280
0.6	5.3280	3.5804	0.7161
0.8	6.0441	-0.2664	-0.0533
1.0	5.9908	0.0551	0.0110



- e. If we apply Euler's method to approximate the solution with $y(0) = 6$, the value of $y_i = 6$ for every value of i since $y = 6$ is an equilibrium solution and the slope of the tangent line will always be 0.

7.3.2 The error in Euler's method

Since we are approximating the solutions to an initial value problem using tangent lines, we should expect that the error in the approximation will be smaller when the step size is smaller. Consider the initial value problem

$$\frac{dy}{dt} = y, \quad y(0) = 1,$$

whose solution we can easily find.

The question posed by this initial value problem is “what function do we know that is the same as its own derivative and has value 1 when $t = 0$?” It is not hard to see that the solution is $y(t) = e^t$. We now apply Euler's method to approximate $y(1) = e$ using several values of Δt . These approximations will be denoted by $E_{\Delta t}$, and we'll use them to see how accurate Euler's Method is.

To begin, we apply Euler's method with a step size of $\Delta t = 0.2$. In that case, we find that $y(1) \approx E_{0.2} = 2.4883$. The error is therefore

$$y(1) - E_{0.2} = e - 2.4883 \approx 0.2300.$$

Repeatedly halving Δt gives the following results, expressed in both tabular and graphical form.

Δt	$E_{\Delta t}$	Error
0.200	2.4883	0.2300
0.100	2.5937	0.1245
0.050	2.6533	0.0650
0.025	2.6851	0.0332

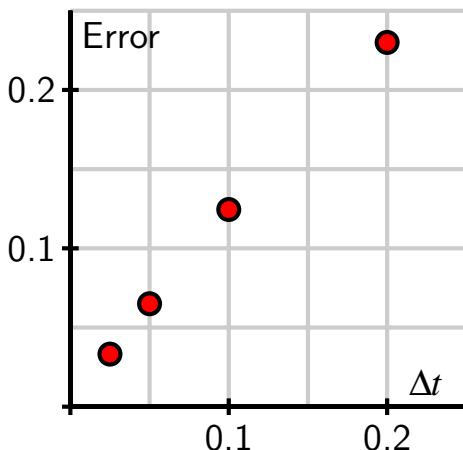


Table 7.3.16: Errors that correspond to different Δt values.

Figure 7.3.17: A plot of the error as a function of Δt .

Notice, both numerically and graphically, that the error is roughly halved when Δt is halved. This example illustrates the following general principle.

If Euler's method is used to approximate the solution to an initial value problem at a point \bar{t} , then the error is proportional to Δt . That is,

$$y(\bar{t}) - E_{\Delta t} \approx K\Delta t$$

for some constant of proportionality K .

7.3.3 Summary

- Euler's method is an algorithm for approximating the solution to an initial value problem by following the tangent lines while we take horizontal steps across the t -axis.
- If we wish to approximate $y(\bar{t})$ for some fixed \bar{t} by taking horizontal steps of size Δt , then the error in our approximation is proportional to Δt .

7.3.4 Exercises

1. (A few steps of Euler's method) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Using Euler's method for a solution of $y' = 4y$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Using Euler's method with different time steps) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.

4. Newton's Law of Cooling says that the rate at which an object, such as a cup of coffee, cools is proportional to the difference in the object's temperature and room temperature. If $T(t)$ is the object's temperature and T_r is room temperature, this law is expressed at

$$\frac{dT}{dt} = -k(T - T_r),$$

where k is a constant of proportionality. In this problem, temperature is measured in degrees Fahrenheit and time in minutes.

- Two calculus students, Alice and Bob, enter a 70° classroom at the same time. Each has a cup of coffee that is 100° . The differential equation for Alice has a constant of proportionality $k = 0.5$, while the constant of proportionality for Bob is $k = 0.1$. What is the initial rate of change for Alice's coffee? What is the initial rate of change for Bob's coffee?
- What feature of Alice's and Bob's cups of coffee could explain this difference?
- As the heating unit turns on and off in the room, the temperature in the room is

$$T_r = 70 + 10 \sin t.$$

Implement Euler's method with a step size of $\Delta t = 0.1$ to approximate the temperature of Alice's coffee over the time interval $0 \leq t \leq 50$. This will most easily be performed using a spreadsheet such as *Excel*. Graph the temperature of her coffee and room temperature over this interval.

- In the same way, implement Euler's method to approximate the temperature of Bob's coffee over the same time interval. Graph the temperature of his coffee and room temperature over the interval.
- Explain the similarities and differences that you see in the behavior of Alice's and Bob's cups of coffee.

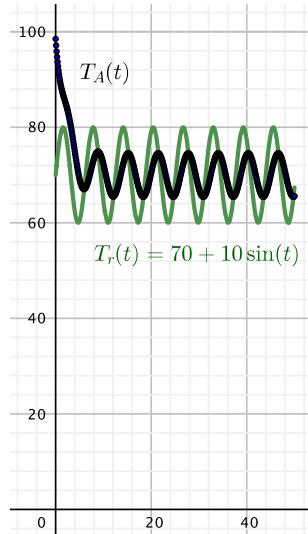
Answer.

- Alice's coffee: $\frac{dT_A}{dt}|_{T=100} = -0.5(30) = -15$ degrees per minute; Bob's coffee: $\frac{dT_B}{dt}|_{T=100} = -0.1(30) = -3$ degrees per minute.
- Consider the insulation of the containers.
- Alice's coffee:

$$\frac{dT_A}{dt} = -0.5(T_A - (70 + 10 \sin t)),$$

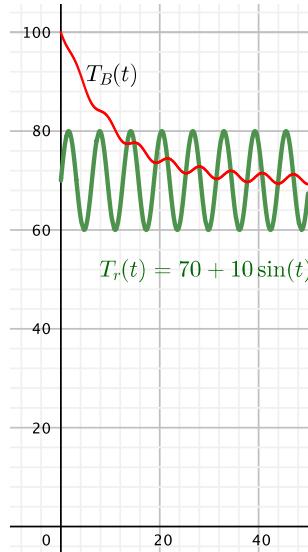
with the initial condition $T_A(0) = 100$.

t	$T_A(t)$
0.0	100
0.1	98.5
0.2	97.12492
0.3	95.86801
0.4	94.72237
\vdots	\vdots
49.6	65.56715
49.7	65.48008
49.8	65.43816
49.9	65.44183
50	65.49103



d.

t	$T_A(t)$
0.0	100
0.1	99.7
0.2	99.41298
0.3	99.13872
0.4	98.87689
\vdots	\vdots
49.6	69.39515
49.7	69.33946
49.8	69.29248
49.9	69.25467
50	69.22638



e. Compare the rate of initial decrease and amplitude of oscillation.

Solution.

a. The differential equation for Alice's coffee is

$$\frac{dT_A}{dt} = -0.5(T - 70),$$

so at the instant $t = 0$, $T = 100$, and thus $\frac{dT_A}{dt}|_{T=100} = -0.5(30) = -15$ degrees per minute. For Bob's coffee,

$$\frac{dT_B}{dt} = -0.1(T - 70),$$

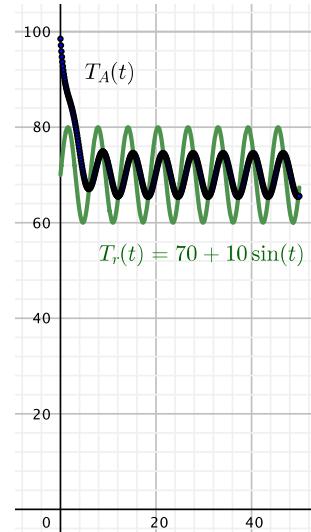
so $\frac{dT_B}{dt}|_{T=100} = -0.1(30) = -3$ degrees per minute. These are the initial rates of change of the respective cups of coffee.

- b. The difference between the cups of coffee could be explained by how insulated the containers are: Bob's coffee loses temperature less quickly, so his is likely better insulated.
- c. With this varying temperature, the governing differential equation for Alice's coffee is

$$\frac{dT_A}{dt} = -0.5(T_A - (70 + 10 \sin t)),$$

with the initial condition $T_A(0) = 100$. Implementing Euler's Method with $\Delta t = 0.1$, we get the following (abbreviated) table of data.

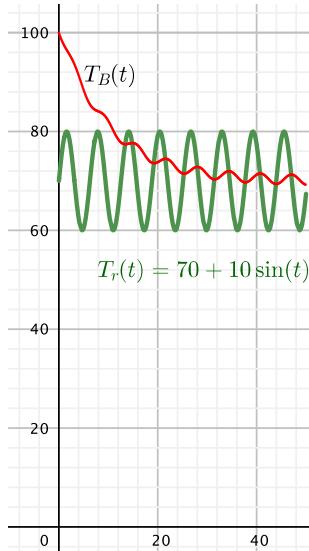
t	$T_A(t)$
0.0	100
0.1	98.5
0.2	97.12492
0.3	95.86801
0.4	94.72237
\vdots	\vdots
49.6	65.56715
49.7	65.48008
49.8	65.43816
49.9	65.44183
50	65.49103



Plotting the data points that result from Euler's Method along with the room temperature function, we see how the room temperature drives the temperature of the coffee up and down as time goes on, as pictured in the figure above, where the data points for the coffee's approximate temperature are shown in dark blue, while the room temperature function, $T_r(t) = 70 + 10 \sin(t)$ is shown in green.

- d. Similar work with the differential equation for Bob's coffee yields the following abbreviated table of approximate function values along with the plot of the data and the temperature function.

t	$T_A(t)$
0.0	100
0.1	99.7
0.2	99.41298
0.3	99.13872
0.4	98.87689
\vdots	\vdots
49.6	69.39515
49.7	69.33946
49.8	69.29248
49.9	69.25467
50	69.22638



- e. The behaviors of both cups of coffee are similar in that each decreases toward the room temperature initially and then is driven in an oscillating way by the room temperature function as it rises and falls. We see that Alice's coffee is more susceptible to being influenced away from the average room temperature, which is consistent with her mug being less well insulated, while in Bob's coffee, we see much less radical oscillations (and a slower overall rate of decrease to get near room temperature), which is consistent with being better insulated.
5. We have seen that the error in approximating the solution to an initial value problem is proportional to Δt . That is, if $E_{\Delta t}$ is the Euler's method approximation to the solution to an initial value problem at \bar{t} , then

$$y(\bar{t}) - E_{\Delta t} \approx K\Delta t$$

for some constant of proportionality K .

In this problem, we will see how to use this fact to improve our estimates, using an idea called *accelerated convergence*.

- a. We will create a new approximation by assuming the error is *exactly* proportional to Δt , according to the formula

$$y(\bar{t}) - E_{\Delta t} = K\Delta t.$$

Using our earlier results from the initial value problem $dy/dt = y$ and $y(0) = 1$ with $\Delta t = 0.2$ and $\Delta t = 0.1$, we have

$$\begin{aligned} y(1) - 2.4883 &= 0.2K \\ y(1) - 2.5937 &= 0.1K. \end{aligned}$$

This is a system of two linear equations in the unknowns $y(1)$ and K . Solve this system to find a new approximation for $y(1)$. (You may remember that the exact value is $y(1) = e = 2.71828\dots$)

- b. Use the other data, $E_{0.05} = 2.6533$ and $E_{0.025} = 2.6851$ to do similar work as in (a) to obtain another approximation. Which gives the better approximation? Why do you think this is?

c. Let's now study the initial value problem

$$\frac{dy}{dt} = t - y, \quad y(0) = 0.$$

Approximate $y(0.3)$ by applying Euler's method to find approximations $E_{0.1}$ and $E_{0.05}$. Now use the idea of accelerated convergence to obtain a better approximation. (For the sake of comparison, you want to note that the actual value is $y(0.3) = 0.0408$.)

Answer.

- a. $K = 1.054$; $y(1) = 2.6991$.
- b. $K = 1.272$; $y(1) = 2.7169$.
- c. $K = 0.122$ and $y(0.3) = 0.0412$.

Solution.

- a. To solve the linear system

$$\begin{aligned} y(1) - 2.4883 &= 0.2K \\ y(1) - 2.5937 &= 0.1K \end{aligned}$$

we subtract the second equation from the first to eliminate $y(1)$. Doing so, we find that

$$-2.4883 - (-2.5937) = 0.2K - 0.1K$$

so that

$$0.1054 = 0.1K$$

and thus $K = 1.054$. Substituting this value into the first equation in the system, it follows that

$$y(1) - 2.4883 = 0.2(1.054)$$

so $y(1) = 0.2(1.054) + 2.4883 = 2.6991$.

- b. Recalling that $E_{0.05} = 2.6533$ and $E_{0.025} = 2.6851$, it follows that this data generates the similar linear system given by

$$\begin{aligned} y(1) - 2.6533 &= 0.05K \\ y(1) - 2.6851 &= 0.025K. \end{aligned}$$

Solving this system yields $K = 1.272$ and $y(1) = 2.7169$. Knowing that the true value of $y(1)$ is $e = 2.71828 \dots$, we see that the approximation here in (b) is much better. That is expected because we are already working with better approximations that come from using $\Delta t = 0.05$ and $\Delta t = 0.025$, which have been twice halved from the smaller Δt value in (a).

- c. To compute $E_{0.1}$ and $E_{0.05}$, we first compute the estimates to $y(0.3)$ using $\Delta t = 0.1$ and $\Delta t = 0.05$. Doing so and making similar assumptions about the error as we did in (a) and (b), we find that for $\Delta t = 0.1$, $y(1) \approx 0.0290$ and for $\Delta t = 0.05$, $y(1) \approx 0.0351$, and therefore we consider the system

$$\begin{aligned} y(0.3) - 0.0290 &= 0.1K \\ y(0.3) - 0.0351 &= 0.05K. \end{aligned}$$

Solving the system, it follows that $K = (-0.0290 + 0.0351)/0.05 = 0.122$ and $y(0.3) = 0.1(0.122) + 0.0290 = 0.0412$, which is much closer to the actual value of $y(0.3) = 0.0408$.

6. In this problem, we'll modify Euler's method to obtain better approximations to solutions of initial value problems. This method is called the *Improved Euler's method*.

In Euler's method, we walk across an interval of width Δt using the slope obtained from the differential equation at the left endpoint of the interval. Of course, the slope of the solution will most likely change over this interval. We can improve our approximation by trying to incorporate the change in the slope over the interval.

Let's again consider the initial value problem $dy/dt = y$ and $y(0) = 1$, which we will approximate using steps of width $\Delta t = 0.2$. Our first interval is therefore $0 \leq t \leq 0.2$. At $t = 0$, the differential equation tells us that the slope is 1, and the approximation we obtain from Euler's method is that $y(0.2) \approx y_1 = 1 + 1(0.2) = 1.2$.

This gives us some idea for how the slope has changed over the interval $0 \leq t \leq 0.2$. We know the slope at $t = 0$ is 1, while the slope at $t = 0.2$ is 1.2, trusting in the Euler's method approximation. We will therefore refine our estimate of the initial slope to be the average of these two slopes; that is, we will estimate the slope to be $(1 + 1.2)/2 = 1.1$. This gives the new approximation $y(1) = y_1 = 1 + 1.1(0.2) = 1.22$.

The first few steps look like what is found in Table 7.3.18.

t_i	y_i	Slope at (t_{i+1}, y_{i+1})	Average slope
0.0	1.0000	1.2000	1.1000
0.2	1.2200	1.4640	1.3420
0.4	1.4884	1.7861	1.6372
:	:	:	:

Table 7.3.18: The first several steps of the improved Euler's method

- Continue with this method to obtain an approximation for $y(1) = e$.
- Repeat this method with $\Delta t = 0.1$ to obtain a better approximation for $y(1)$.
- We saw that the error in Euler's method is proportional to Δt . Using your results from parts (a) and (b), what power of Δt appears to be proportional to the error in the Improved Euler's Method?

Answer.

- $y(1) \approx y_5 = 2.7027$.
- $y(1) \approx y_{10} = 2.7141$.
- The square of Δt .

Solution.

- Continuing with this method, we find $y(0.6) \approx y_3 = 1.8156$, $y(0.8) \approx y_4 = 2.2153$, and $y(1) \approx y_5 = 2.7027$, which is an approximation for $y(1) = e$.
- Updating this method with $\Delta t = 0.1$, we find $y(0.1) \approx y_1 = 1.105$, $y(0.2) \approx y_2 = 1.221$, ..., $y(1) \approx y_{10} = 2.7141$, which is an even better approximation for $y(1) = e$.
- We have halved the value of Δt in these two approximations of $y(1)$ in the Improved Euler Method. Using the notation $I_{0.2} = 2.7027$ and $I_{0.1} = 2.7141$ for the estimates

generated by the method with the respective t -values, we see that the following occurs with the error (using the known value of $y(1) = e$):

$$y(1) - I_{0.2} = 2.7183 - 2.7027 = 0.0156$$

$$y(1) - I_{0.1} = 2.7183 - 2.7141 = 0.0042$$

Computing $I_{0.05}$, we find $I_{0.05} = 2.7172$, and thus $y(1) - I_{0.05} = 0.0011$. Experimenting with powers of Δt , it appears that these errors are each approximately $0.4\Delta t^2$, and thus the error in the Improved Euler Method seems to be proportional to the square of Δt .

7.4 Separable differential equations

Motivating Questions

- What is a separable differential equation?
- How can we find solutions to a separable differential equation?
- Are some of the differential equations that arise in applications separable?

In Sections 7.2 and 7.3, we have seen several ways to approximate the solution to an initial value problem. Given the frequency with which differential equations arise in the world around us, we would like to have some techniques for finding explicit algebraic solutions of certain initial value problems. In this section, we focus on a particular class of differential equations (called *separable*) and develop a method for finding algebraic formulas for their solutions.

A *separable differential equation* is a differential equation whose algebraic structure allows the variables to be separated in a particular way. For instance, consider the equation

$$\frac{dy}{dt} = ty.$$

We would like to separate the variables t and y so that all occurrences of t appear on the right-hand side, and all occurrences of y appear on the left, multiplied by dy/dt . For this example, we divide both sides by y so that

$$\frac{1}{y} \frac{dy}{dt} = t.$$

Note that when we attempt to separate the variables in a differential equation, we require that one side is a product in which the derivative dy/dt is one factor and the other factor is solely an expression involving y .

Not every differential equation is separable. For example, if we consider the equation

$$\frac{dy}{dt} = t - y,$$

it may seem natural to separate it by writing

$$y + \frac{dy}{dt} = t.$$

As we will see, this will not be helpful, since the left-hand side is not a product of a function of y with $\frac{dy}{dt}$.

Example 7.4.1. In this example, we explore whether certain differential equations are separable or not, and then revisit some key ideas from earlier work in integral calculus.

- a. Which of the following differential equations are separable? If the equation is separable, write the equation in the revised form $g(y) \frac{dy}{dt} = h(t)$.

- i. $\frac{dy}{dt} = -3y$.
- ii. $\frac{dy}{dt} = ty - y$.
- iii. $\frac{dy}{dt} = t + 1$.
- iv. $\frac{dy}{dt} = t^2 - y^2$.

- b. Explain why any autonomous differential equation is guaranteed to be separable.
c. Why do we include the term “ $+C$ ” in the expression

$$\int x \, dx = \frac{x^2}{2} + C?$$

- d. Suppose we know that a certain function f satisfies the equation

$$\int f'(x) \, dx = \int x \, dx.$$

What can you conclude about f ?

7.4.1 Solving separable differential equations

Before we discuss a general approach to solving a separable differential equation, it is instructive to consider an example.

Example 7.4.2. Find all functions y that are solutions to the differential equation

$$\frac{dy}{dt} = \frac{t}{y^2}.$$

Solution. We begin by separating the variables and writing

$$y^2 \frac{dy}{dt} = t.$$

Integrating both sides of the equation with respect to the independent variable t shows that

$$\int y^2 \frac{dy}{dt} \, dt = \int t \, dt.$$

Next, we notice that the left-hand side allows us to change the variable of antiderivatiation¹ from t to y . In particular, $dy = \frac{dy}{dt} dt$, so we now have

$$\int y^2 dy = \int t dt.$$

This equation says that two families of antiderivatives are equal to each other. Therefore, when we find representative antiderivatives of both sides, we know they must differ by an arbitrary constant C . Antiderivatiation and including the integration constant C on the right, we find that

$$\frac{y^3}{3} = \frac{t^2}{2} + C.$$

It is not necessary to include an arbitrary constant on both sides of the equation; we know that $y^3/3$ and $t^2/2$ are in the same family of antiderivatives and must therefore differ by a single constant.

Finally, we solve the last equation above for y as a function of t , which gives

$$y(t) = \sqrt[3]{\frac{3}{2} t^2 + 3C}.$$

Of course, the term $3C$ on the right-hand side represents 3 times an unknown constant. It is, therefore, still an unknown constant, which we will rewrite as C . We thus conclude that the function

$$y(t) = \sqrt[3]{\frac{3}{2} t^2 + C}$$

is a solution to the original differential equation for any value of C .

Notice that because this solution depends on the arbitrary constant C , we have found an infinite family of solutions. This makes sense because we expect to find a unique solution that corresponds to any given initial value.

For example, if we want to solve the initial value problem

$$\frac{dy}{dt} = \frac{t}{y^2}, \quad y(0) = 2,$$

we know that the solution has the form $y(t) = \sqrt[3]{\frac{3}{2} t^2 + C}$ for some constant C . We therefore must find the appropriate value for C that gives the initial value $y(0) = 2$. Hence,

$$2 = y(0) = \sqrt[3]{\frac{3}{2} 0^2 + C} = \sqrt[3]{C},$$

which shows that $C = 2^3 = 8$. The solution to the initial value problem is then

$$y(t) = \sqrt[3]{\frac{3}{2} t^2 + 8}.$$

¹This is why we required that the left-hand side be written as a product in which dy/dt is one of the terms.

The strategy of Example 7.4.2 may be applied to any differential equation of the form $\frac{dy}{dt} = g(y) \cdot h(t)$, and any differential equation of this form is said to be *separable*. We work to solve a separable differential equation by writing

$$\frac{1}{g(y)} \frac{dy}{dt} = h(t),$$

and then integrating both sides with respect to t . After integrating, we try to solve algebraically for y in order to write y as a function of t .

Example 7.4.3. Solve the differential equation

$$\frac{dy}{dt} = 3y.$$

Solution. Following the same strategy as in Example 7.4.2, we have

$$\frac{1}{y} \frac{dy}{dt} = 3.$$

Integrating both sides with respect to t ,

$$\int \frac{1}{y} \frac{dy}{dt} dt = \int 3 dt,$$

and thus

$$\int \frac{1}{y} dy = \int 3 dt.$$

Antidifferentiating and including the integration constant, we find that

$$\ln |y| = 3t + C.$$

Finally, we need to solve for y . Here, one point deserves careful attention. By the definition of the natural logarithm function, it follows that

$$|y| = e^{3t+C} = e^{3t} e^C.$$

Since C is an unknown constant, e^C is as well, though we do know that it is positive (because e^x is positive for any x). When we remove the absolute value in order to solve for y , however, this constant may be either positive or negative. To account for a possible $+$ or $-$, we denote this updated constant by C to obtain

$$y(t) = C e^{3t}.$$

There is one more technical point to make. Notice that $y = 0$ is an equilibrium solution to this differential equation. In solving the equation above, we begin by dividing both sides by y , which is not allowed if $y = 0$. To be perfectly careful, therefore, we should consider the equilibrium solutions separately. In this case, notice that the final form of our solution captures the equilibrium solution by allowing $C = 0$.

Example 7.4.4. Suppose that the population of a town is growing continuously at an annual rate of 3% per year.

- Let $P(t)$ be the population of the town in year t . Write a differential equation that describes the annual growth rate.
- Find the solutions of this differential equation.
- If you know that the town's population in year 0 is 10,000, find the population $P(t)$.
- How long does it take for the population to double? This time is called the *doubling time*.
- Working more generally, find the doubling time if the annual growth rate is k times the population.

Hint.

- Small hints for each of the prompts above.

Answer.

- $\frac{dP}{dt} = 0.03P$
- $P = Ce^{0.03t}$.
- $P = 10000e^{0.03t}$.
- The doubling time is $t = \frac{\ln(2)}{0.03} \approx 23.105$ years.
- The doubling time is $t = \frac{1}{k} \ln(2)$.

Solution.

- Since the population $P(t)$ of the town is growing at an annual rate of 3% per year, this means that $\frac{dP}{dt} = 0.03P$
- We use separation of variables and write $\frac{1}{P} \frac{dP}{dt} = 0.03$. We then have

$$\int \frac{1}{P} \frac{dP}{dt} dt = \int 0.03 dt$$

and thus $\int \frac{1}{P} dP = \int 0.03 dt$. Evaluating the integrals on each side, we have $\ln |P| = 0.03t + c$. By definition of the logarithm function, it follows that

$$|P| = e^{0.03t+c} = e^c e^{0.03t} = Ce^{0.03t}.$$

Incorporating the \pm that follows from the absolute value sign into the constant C , we conclude that $P = Ce^{0.03t}$.

- c. Given that $P(0) = 10000$, we see that $10000 = Ce^0 = C$. So

$$P = 10000e^{0.03t}.$$

- d. The population will have doubled when $P = 200000$. To determine the time it takes the population to double, we solve the equation

$$10000e^{0.03t} = 20000.$$

Thus $e^{0.03t} = 2$, so $t = \frac{\ln(2)}{0.03} \approx 23.105$. It will take about 23.1 years for the population to double in size.

- e. If the annual growth rate is k , then the differential equation for the population is

$$\frac{dP}{dt} = kP.$$

We can use separation of variables to show that $P = P_0e^{kt}$, where P_0 is the population at time $t = 0$. We then find the doubling time by solving the equation $P_0e^{kt} = 2P_0$. We first solve for e^{kt} and get $e^{kt} = 2$. Taking natural logs, $kt = \ln(2)$, and dividing by k , $t = \frac{1}{k} \ln(2)$. The doubling time for the population is thus $\frac{1}{k} \ln(2)$.

Example 7.4.5. Suppose that a cup of coffee is initially at a temperature of 105° F and is placed in a 75° F room. Newton's law of cooling says that

$$\frac{dT}{dt} = -k(T - 75),$$

where k is a constant of proportionality.

- Suppose you measure that the coffee is cooling at one degree per minute at the time the coffee is brought into the room. Use the differential equation to determine the value of the constant k .
- Find all the solutions of this differential equation.
- What happens to all the solutions as $t \rightarrow \infty$? Explain how this agrees with your intuition.
- What is the temperature of the cup of coffee after 20 minutes?
- How long does it take for the coffee to cool to 80° ?

Hint.

- Small hints for each of the prompts above.

Answer.

- $k = \frac{1}{30}$

- b. $T = 75 + Ce^{-t/30}$
- c. The temperature of the coffee tends to 75 degrees.
- d. $T(20) = 75 + 30e^{-2/3} \approx 90.4^\circ\text{F}$.
- e. $t = -30 \ln\left(\frac{1}{6}\right) \approx 53.75$ minutes.

Solution. We have

$$\frac{dT}{dt} = -k(T - 75),$$

where k is a constant. We also have $T(0) = 105$.

- a. Since the coffee is cooling at a rate of one degree per minute at time 0, it follows that $\frac{dT}{dt} = -1$ at this instant. From the given differential equation, we know $-k(T(0) - 75) = -1$. Next, $T(0) = 105$, so $-k(105 - 75) = -1$, and thus $k = \frac{1}{30}$.
- b. We use separation of variables to solve the differential equation $\frac{dT}{dt} = -\frac{1}{30}(T - 75)$. Separating variables and integrating both sides with respect to t ,

$$\int \frac{1}{T - 75} dT = - \int \frac{1}{30} dt$$

Integrating both sides,

$$\ln|T - 75| = -\frac{1}{30}t + c$$

Using the definition of the logarithm function and writing $e^c = C$, we see that

$$|T - 75| = e^{-t/30+c} = Ce^{-t/30}$$

Finally, including the absolute value in C and solving for T , $T - 75 = Ce^{-t/30}$, so $T = 75 + Ce^{-t/30}$.

- c. As $t \rightarrow \infty$, $e^{-t/30} \rightarrow 0$. So

$$\lim_{t \rightarrow \infty} (75 + Ce^{-t/30}) = 75.$$

This means that as $t \rightarrow \infty$, $T \rightarrow 75$, or the temperature of the coffee tends to 75 degrees.

- d. In order to determine the temperature of the coffee after 20 minutes, we need to first determine the value of the constant C in the solution $T = 75 + Ce^{-t/30}$. To do this, we use the fact that the temperature of the coffee at $t = 0$ is 105°F . Thus, $105 = 75 + Ce^0$, so $C = 30$. Now we know that $T = 75 + 30e^{-t/30}$, and when $t = 20$, $T(20) = 75 + 30e^{-2/3}$. The temperature of the coffee after 20 minutes is approximately 90.4°F .
- e. To determine how long it will take the coffee to cool to 80°F , we solve the equation $80 = 75 + 30e^{-t/30}$ for t . Solving first for $e^{-t/30}$, we subtract 75 from both sides and divide by 30 to get $e^{-t/30} = \frac{5}{30}$. Taking logarithms and solving for t , $-\frac{t}{30} = \ln\left(\frac{1}{6}\right)$, and thus $t = -30 \ln\left(\frac{1}{6}\right) \approx 53.75$ minutes. We conclude that it will take about 53.75 minutes for the coffee to cool to 80°F .

Example 7.4.6. Solve each of the following differential equations or initial value problems.

- a. $\frac{dy}{dt} - (2-t)y = 2-t$
 b. $\frac{1}{t} \frac{dy}{dt} = e^{t^2-2y}$
 c. $y' = 2y + 2, y(0) = 2$
 d. $y' = 2y^2, y(-1) = 2$
 e. $\frac{dy}{dt} = \frac{-2ty}{t^2+1}, y(0) = 4$

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a. $y = -1 + Ce^{\left(2t - \frac{t^2}{2}\right)}$.
 b. $y = \frac{1}{2} \ln \left(e^{t^2} + C \right)$.
 c. $y = -1 + 3e^{2t}$.
 d. $y = -\frac{1}{2t+\frac{3}{2}} = -\frac{2}{4t+3}$.
 e. $y = \frac{4}{t^2+1}$.

Solution.

- a. We are given that $\frac{dy}{dt} - (2-t)y = 2-t$. We first separate the variables by writing $\frac{dy}{dt} = (2-t)y + (2-t) = (2-t)(y+1)$ and then dividing both sides by $(y+1)$ to find

$$\frac{1}{1+y} \frac{dy}{dt} = 2-t.$$

Integrating both sides with respect to t , we have

$$\int \frac{1}{1+y} dy = \int (2-t) dt.$$

Evaluating the integrals,

$$\ln|1+y| = \left(2t - \frac{1}{2}t^2\right) + c$$

Rewriting without logarithms,

$$|1+y| = e^{\left(2t - \frac{t^2}{2}\right)} e^c$$

Writing $e^c = C$ and including the sign from the absolute value in this constant,

$$1+y = Ce^{\left(2t - \frac{t^2}{2}\right)},$$

and therefore $y = -1 + Ce^{\left(2t - \frac{t^2}{2}\right)}$.

b. Given that $\frac{1}{t} \frac{dy}{dt} = e^{t^2-2y}$, we first observe that $e^{t^2-2y} = e^{t^2} e^{-2y}$. Separating the variables,

$$e^{2y} \frac{dy}{dt} = t e^{t^2}.$$

Integrating with respect to t ,

$$\int e^{2y} dy = \int t e^{t^2} dt.$$

Using standard substitution techniques on both sides, we find that

$$\frac{1}{2} e^{2y} = \frac{1}{2} e^{t^2} + c$$

Writing $2c = C$, we then have $e^{2y} = e^{t^2} + C$, and taking logarithms to solve for y , we find $2y = \ln(e^{t^2} + C)$, and thus

$$y = \frac{1}{2} \ln(e^{t^2} + C).$$

c. For the initial value problem $y' = 2y + 2$, $y(0) = 2$, we first solve the differential equation. Noting that $2y + 2 = 2(y + 1)$, we have

$$\frac{dy}{dt} = 2(y + 1),$$

so

$$\frac{1}{y+1} \frac{dy}{dt} = 2$$

Integrating with respect to t ,

$$\int \frac{1}{y+1} dy = \int 2 dt$$

and thus evaluating the integrals, $\ln|y+1| = 2t + C$. It follows that $|y+1| = e^{2t+C} = e^C e^{2t}$, so in the usual way we can write $y+1 = C e^{2t}$, and therefore

$$y = -1 + C e^{2t}.$$

Using $y(0) = 2$ to solve the IVP, we obtain $2 = -1 + C e^0$ and so $C = 3$. The solution to the initial value problem is

$$y = -1 + 3e^{2t}.$$

d. For the initial value problem $y' = 2y^2$, $y(-1) = 2$, we first separate variables and write

$$\frac{1}{y^2} \frac{dy}{dt} = 2.$$

Next, integrating with respect to t ,

$$\int \frac{1}{y^2} dy = \int 2 dt,$$

and thus $-y^{-1} = 2t + c$. Solving for y , it follows

$$y = -\frac{1}{2t + c}.$$

We now use the initial value $y(-1) = 2$ and obtain $2 = -\frac{1}{-2+c}$. Solving for c gives $c = \frac{3}{2}$. The solution of the initial value problem is thus

$$y = -\frac{1}{2t + \frac{3}{2}} = -\frac{2}{4t + 3}.$$

e. For the IVP $\frac{dy}{dt} = \frac{-2ty}{t^2+1}$, $y(0) = 4$, we first separate variables and write

$$\frac{1}{y} \frac{dy}{dt} = \frac{-2t}{t^2+1}.$$

Integrating with respect to t we find

$$\int \frac{1}{y} dy = - \int \frac{2t}{t^2+1} dt.$$

Using the substitution $u = t^2 + 1$ on the right, it follows that

$$\ln|y| = -\ln|t^2 + 1| + c.$$

By definition and properties of the logarithm,

$$|y| = e^{-\ln|t^2+1|} e^c = e^{\ln|t^2+1|^{-1}} e^c = |t^2+1|^{-1} e^c.$$

Writing $C = e^c$ and absorbing the absolute value from the left side, we have

$$y = \frac{C}{t^2+1}.$$

Using the initial condition $y(0) = 4$, we obtain $4 = \frac{C}{1}$ or $C = 4$. Thus, the solution to the initial value problem is

$$y = \frac{4}{t^2+1}.$$

7.4.2 Summary

- A separable differential equation is one that may be rewritten with all occurrences of the dependent variable multiplying the derivative and all occurrences of the independent variable on the other side of the equation.
- We may find the solutions to certain separable differential equations by separating variables, integrating with respect to t , and ultimately solving the resulting algebraic equation for y .
- This technique allows us to solve many important differential equations that arise in the world around us. For instance, questions of growth and decay and Newton's Law of Cooling give rise to separable differential equations. Later, we will learn in Section 7.6 that the important logistic differential equation is also separable.

7.4.3 Exercises

- (Initial value problem for $dy/dx = x^8y$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Initial value problem for $dy/dt = 0.9(y - 300)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Initial value problem for $dy/dt = y^2(8 + t)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Initial value problem for $du/dt = e^{6u+10t}$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- (Initial value problem for $dy/dx = 170yx^{16}$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
- The mass of a radioactive sample decays at a rate that is proportional to its mass.
 - Express this fact as a differential equation for the mass $M(t)$ using k for the constant of proportionality.
 - If the initial mass is M_0 , find an expression for the mass $M(t)$.
 - The *half-life* of the sample is the amount of time required for half of the mass to decay. Knowing that the half-life of Carbon-14 is 5730 years, find the value of k for a sample of Carbon-14.
 - How long does it take for a sample of Carbon-14 to be reduced to one-quarter its original mass?
 - Carbon-14 naturally occurs in our environment; any living organism takes in Carbon-14 when it eats and breathes. Upon dying, however, the organism no longer takes in Carbon-14. Suppose that you find remnants of a pre-historic firepit. By analyzing the charred wood in the pit, you determine that the amount of Carbon-14 is only 30% of the amount in living trees. Estimate the age of the firepit.²

Answer.

- $\frac{dM}{dt} = kM$.
- $M(t) = M_0 e^{kt}$.
- $M(t) = M_0 e^{-\frac{\ln(2)}{5730}t} \approx M_0 e^{-0.000121t}$
- $t = \frac{5730 \ln(4)}{\ln(2)} \approx 11460$ years.
- $t = -\frac{5730 \ln(0.3)}{\ln(2)} \approx 9952.8$ years.

Solution.

- A differential equation that represents the given description is

$$\frac{dM}{dt} = kM.$$

²This approach is the basic idea behind radiocarbon dating.

- b. Separating the variables to solve the differential equation yields

$$\frac{1}{M} dM = k dt$$

so

$$\int \frac{1}{M} dM = \int k dt.$$

Integrating and solving for M , we find $\ln|M| = kt + C$ and thus $M = e^{kt+C} = e^{kt}e^C = Ke^{kt}$ for some constant K . Using the initial condition $M(0) = M_0$, it follows that $K = M_0$. Therefore, the mass at time t is given by

$$M(t) = M_0 e^{kt}.$$

- c. If we let 5730 years pass from time 0, the remaining mass of the radioactive sample will be $\frac{1}{2}M_0$. This allows us to find the value of k by setting $M(5730)$ equal to $\frac{1}{2}M_0$ and solving for t . Doing so, we have $\frac{1}{2}M_0 = M_0 e^{5730k}$, and thus $\frac{1}{2} = e^{5730k}$. It follows that

$$\ln\left(\frac{1}{2}\right) = 5730k$$

from which we find $k = \frac{1}{5730} \ln\left(\frac{1}{2}\right) = -\frac{\ln(2)}{5730}$. Hence,

$$M(t) = M_0 e^{-\frac{\ln(2)}{5730}t} \approx M_0 e^{-0.000121t}$$

- d. Now we want to find the time t so that $M(t) = \frac{1}{4}M_0$. Similar work to our recent efforts in (c) shows that $t = \frac{5730 \ln(4)}{\ln(2)} \approx 11460$ years. This number makes sense because it takes 5730 years for half of the sample to decay, then another 5730 years for half to decay again, for a total of $(2)(5730)$ years for the substance to decay to one quarter of its original amount.
- e. We want to find the time it takes for $M(t)$ to be 30% of M_0 , so we need to solve $0.3M_0 = M_0 e^{-\frac{\ln(2)}{5730}t}$. Doing so in the usual way, we find that $t = -\frac{5730 \ln(0.3)}{\ln(2)} \approx 9952.8$ years, so we estimate that the sample is about 9973 years old.

7. Consider the initial value problem

$$\frac{dy}{dt} = -\frac{t}{y}, \quad y(0) = 8$$

- a. Find the solution of the initial value problem and sketch its graph.
 b. For what values of t is the solution defined?
 c. What is the value of y at the last time that the solution is defined?
 d. By looking at the differential equation, explain why we should not expect to find solutions with the value of y you noted in (c).

Answer.

- a. $y = \sqrt{64 - t^2}$.
 b. $-8 \leq t \leq 8$.

- c. $y(8) = 0$.
- d. $\frac{dy}{dt} = -\frac{t}{y}$ is not defined when $y = 0$.

Solution.

- a. Separating the variables, we find that

$$y \frac{dy}{dt} = -t$$

and integrating with respect to t

$$\int y \, dy = - \int t \, dt.$$

Evaluating the integrals and solving for y , we see that $\frac{1}{2}y^2 = -\frac{1}{2}t^2 + C$ and therefore

$$y^2 = C - t^2,$$

so $y = \pm\sqrt{C - t^2}$. Applying the initial condition $y(0) = 8$, we find that $8 = \sqrt{C}$ (we choose the $+$ from \pm since $8 > 0$), so $C = 64$. Hence the solution to the IVP is

$$y = \sqrt{64 - t^2}.$$

- b. The solution to the IVP is defined for $-8 \leq t \leq 8$.
 - c. The last time that the solution is defined is $t = 8$ and the corresponding value of y at this point is $y(8) = 0$.
 - d. Recall that the differential equation is $\frac{dy}{dt} = -\frac{t}{y}$, and the righthand expression is undefined when $y = 0$. Hence the solution value is technically not defined at the point we noted in (c); this is connected to there being a vertical tangent line at this point in the direction field.
8. Suppose that a cylindrical water tank with a hole in the bottom is filled with water. The water, of course, will leak out and the height of the water will decrease. Let $h(t)$ denote the height of the water. A physical principle called *Torricelli's Law* implies that the height decreases at a rate proportional to the square root of the height.
- a. Express this fact using k as the constant of proportionality.
 - b. Suppose you have two tanks, one with $k = -1$ and another with $k = -10$. What physical differences would you expect to find?
 - c. Suppose you have a tank for which the height decreases at 20 inches per minute when the water is filled to a depth of 100 inches. Find the value of k .
 - d. Solve the initial value problem for the tank in part (c), and graph the solution you determine.
 - e. How long does it take for the water to run out of the tank?
 - f. Is the solution that you found valid for all time t ? If so, explain how you know this. If not, explain why not.

Answer.

- a. $\frac{dh}{dt} = k\sqrt{h}$.
- b. The tank with $k = -10$ has water leaving the tank much more rapidly.
- c. $k = -5$.
- d. $h(t) = (5 - 2.5t)^2$.
- e. 2.5 minutes.
- f. No.

Solution.

- a. Since $\frac{dh}{dt}$ is proportional to the square root of the height of the water, we have

$$\frac{dh}{dt} = k\sqrt{h}.$$

- b. If one tank has $k = -1$ and the other has $k = -10$, the tank with $k = -10$ will have water leaving the tank much more rapidly, which should correspond to that tank having a larger hole.
- c. Since height decreases at 20 inches per minute when the water is filled to a depth of 100 inches, this tells us that $\frac{dh}{dt}|_{h=100} = -20$. Applying this information to the differential equation,

$$-20 = k\sqrt{100}$$

and thus $k = -5$.

- d. To solve the differential equation $\frac{dh}{dt} = -5\sqrt{h}$ with initial condition $h(0) = 100$, we first separate the variables and write

$$\frac{1}{\sqrt{h}} \frac{dh}{dt} = -5.$$

Rewriting the first fraction and integrating with respect to t , we have

$$\int h^{-1/2} dh = \int -5 dt$$

and thus

$$2h^{1/2} = -5t + C,$$

or $h = (C - 2.5t)^2$. Knowing that $h(0) = 100$, we see that $100 = C^2$, and thus $C = 5$. This shows that the solution to the IVP is

$$h(t) = (5 - 2.5t)^2.$$

A plot of this function shows it to be a concave up parabola with y -intercept $(0, 25)$ and a repeated zero at $(2.5, 0)$.

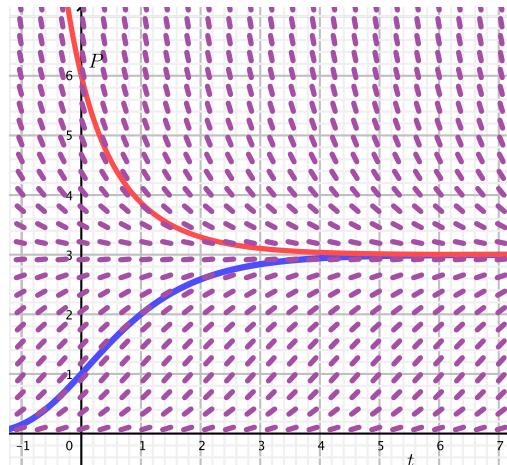
- e. It takes 2.5 minutes for the tank to empty since $h(2.5) = 0$.
 - f. No, this solution is not valid for all time t . The physical constraints of the problem tell us that the solution is only valid for $0 < t < 2.5$ since at $t = 2.5$ the tank is empty.
9. The *Compertz equation* is a model that is used to describe the growth of certain populations. Suppose that $P(t)$ is the population of some organism and that

$$\frac{dP}{dt} = -P \ln\left(\frac{P}{3}\right) = -P(\ln P - \ln 3).$$

- Sketch a slope field for $P(t)$ over the range $0 \leq P \leq 6$.
- Identify any equilibrium solutions and determine whether they are stable or unstable.
- Find the population $P(t)$ assuming that $P(0) = 1$ and sketch its graph. What happens to $P(t)$ after a very long time?
- Find the population $P(t)$ assuming that $P(0) = 6$ and sketch its graph. What happens to $P(t)$ after a very long time?
- Verify that the long-term behavior of your solutions agrees with what you predicted by looking at the slope field.

Answer.

(a)



(b) $P = 3$ is stable.

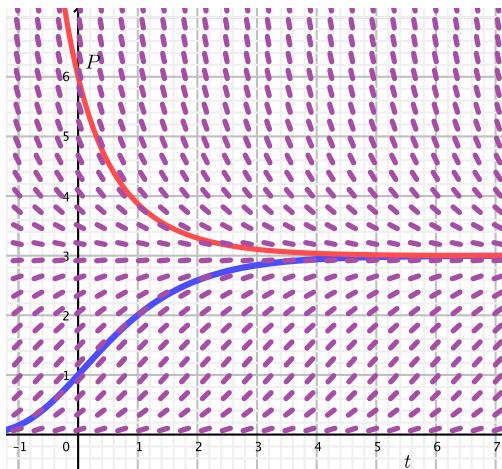
(c) $P(t) = 3e^{\ln(\frac{1}{3})e^{-t}}$.

(d) $P(t) = 3e^{\ln(2)e^{-t}}$.

(e) Yes.

Solution.

(a)



- (b) The only equilibrium solution is $P = 3$, since $\ln(1) = 0$ (note: $\ln(0)$ is undefined, so the differential equation is not defined for $P = 0$). From the slope field pictured in (a), the equilibrium solution $P = 3$ is stable.
- (c) To solve the IVP with $\frac{dP}{dt} = -P \ln\left(\frac{P}{3}\right)$ and $P(0) = 1$, we use separation of variables. Observe that

$$\frac{1}{P \ln\left(\frac{P}{3}\right)} \frac{dP}{dt} = -1$$

and thus

$$\int \frac{1}{P \ln\left(\frac{P}{3}\right)} dP = - \int 1 dt$$

On the left, using the substitution $u = \ln(P/3)$, it follows $du = \frac{1}{P} \cdot \frac{1}{3} dP$, we can rewrite the most recent equation as

$$\int \frac{du}{u} = - \int 1 dt.$$

Therefore,

$$\ln|u| = -t + C,$$

so $|u| = e^{C-t}$, or $u = Ke^{-t}$. Next, we recall that $u = \ln\left(\frac{P}{3}\right)$, so

$$\frac{P}{3} = e^{Ke^{-t}}$$

and $P = 3e^{Ke^{-t}}$. Applying the initial condition, $1 = 3e^K$, so $K = \ln\left(\frac{1}{3}\right)$, and thus

$$P(t) = 3e^{\ln\left(\frac{1}{3}\right)e^{-t}}.$$

This function's plot is shown in blue in the image above in (a) and we see (both graphically and algebraically) that $P(t) \rightarrow 3$ as $t \rightarrow \infty$ since $e^{-t} \rightarrow 0$.

- (d) Changing the initial condition to $P(0) = 6$, we can use all of our preceding work in (c) up to where we established $P = 3e^{Ke^{-t}}$. Applying this initial condition, $6 = 3e^K$, so $K = \ln(2)$, and thus

$$P(t) = 3e^{\ln(2)e^{-t}}.$$

This function's plot is shown in red in the image above in (a) and we see (both graphically and algebraically) that $P(t) \rightarrow 3$ as $t \rightarrow \infty$ since $e^{-t} \rightarrow 0$.

- (e) Our work in (c) and (d) suggests that regardless of the initial condition with $P(0) > 0$, the long-term behavior of P will be for it to tend to 3 as t increases without bound. This matches our initial observations based on the slope field.

7.5 Modeling with differential equations

Motivating Questions

- How can we use differential equations to describe phenomena in the world around us?
 - How can we use differential equations to better understand these phenomena?
-

We have seen several ways that differential equations arise in the natural world, from the growth of a population to the temperature of a cup of coffee. In this section, we look more closely at how differential equations give us a natural way to describe various phenomena. As we'll see, the key is to understand the different factors that cause a quantity to change.

Example 7.5.1. Any time that the rate of change of a quantity is related to the amount of a quantity, a differential equation naturally arises. In the following two problems, we see two such scenarios; for each, we want to develop a differential equation whose solution is the quantity of interest.

- Suppose you have a bank account in which money grows at an annual rate of 3%.
 - If you have \$10,000 in the account, at what rate is your money growing?
 - Suppose that you are also withdrawing money from the account at \$1,000 per year. What is the rate of change in the amount of money in the account? What are the units of this rate of change?
- Suppose that a water tank holds 100 gallons and that a salty solution, which contains 20 grams of salt in every gallon, enters the tank at 2 gallons per minute.
 - How much salt enters the tank each minute?
 - Suppose that initially there are 300 grams of salt in the tank. How much salt is in each gallon at this point in time?
 - Finally, suppose that evenly mixed solution is pumped out of the tank at the rate of 2 gallons per minute. How much salt leaves the tank each minute?
 - What is the total rate of change in the amount of salt in the tank?

7.5.1 Developing a differential equation

Example 7.5.1 demonstrates the kind of thinking we will be doing in this section. In each of the two examples we considered, there is a quantity, such as the amount of money in the bank account or the amount of salt in the tank, that is changing due to several factors. The governing differential equation states that the total rate of change is the difference between the rate of increase and the rate of decrease.

Example 7.5.2. In the Great Lakes region, rivers flowing into the lakes carry a great deal of pollution in the form of small pieces of plastic averaging 1 millimeter in diameter. In order to understand how the amount of plastic in Lake Michigan is changing, construct a model for how this type pollution has built up in the lake.

Solution. First, some basic facts about Lake Michigan.

- The volume of the lake is $5 \cdot 10^{12}$ cubic meters.
- Water flows into the lake at a rate of $5 \cdot 10^{10}$ cubic meters per year. It flows out of the lake at the same rate.
- Each cubic meter flowing into the lake contains roughly $3 \cdot 10^{-8}$ cubic meters of plastic pollution.

Let's denote the amount of pollution in the lake by $P(t)$, where P is measured in cubic meters of plastic and t in years. Our goal is to describe the rate of change of this function; so we want to develop a differential equation describing $P(t)$.

First, we will measure how $P(t)$ increases due to pollution flowing into the lake. We know that $5 \cdot 10^{10}$ cubic meters of water enters the lake every year and each cubic meter of water contains $3 \cdot 10^{-8}$ cubic meters of pollution. Therefore, pollution enters the lake at the rate of

$$\left(5 \cdot 10^{10} \frac{\text{m}^3 \text{ water}}{\text{year}}\right) \left(3 \cdot 10^{-8} \frac{\text{m}^3 \text{ plastic}}{\text{m}^3 \text{ water}}\right) = 1.5 \cdot 10^3 \text{ cubic m of plastic per year.}$$

Second, we will measure how $P(t)$ decreases due to pollution flowing out of the lake. If the total amount of pollution is P cubic meters and the volume of Lake Michigan is $5 \cdot 10^{12}$ cubic meters, then the concentration of plastic pollution in Lake Michigan is

$$\frac{P}{5 \cdot 10^{12}} \text{ cubic m of plastic per cubic m of water.}$$

Since $5 \cdot 10^{10}$ cubic meters of water flow out each year¹, then the plastic pollution leaves the lake at the rate of

$$\left(\frac{P}{5 \cdot 10^{12}} \frac{\text{m}^3 \text{ plastic}}{\text{m}^3 \text{ water}}\right) \left(5 \cdot 10^{10} \frac{\text{m}^3 \text{ water}}{\text{year}}\right) = \frac{P}{100} \text{ cubic m of plastic per year.}$$

The total rate of change of P is thus the difference between the rate at which pollution enters the lake and the rate at which pollution leaves the lake; that is,

$$\begin{aligned} \frac{dP}{dt} &= 1.5 \cdot 10^3 - \frac{P}{100} \\ &= \frac{1}{100}(1.5 \cdot 10^5 - P). \end{aligned}$$

¹and we assume that each cubic meter of water that flows out carries with it the plastic pollution it contains

We have now found a differential equation that describes the rate at which the amount of pollution is changing. To understand the behavior of $P(t)$, we apply some of the techniques we have recently developed.

Because the differential equation is autonomous, we can sketch dP/dt as a function of P and then construct a slope field, as shown in Figure 7.5.3 and Figure 7.5.4.

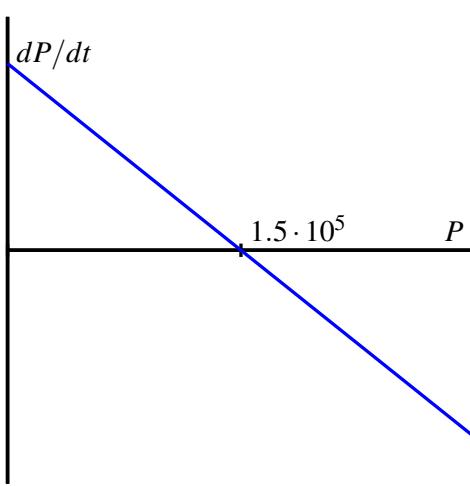


Figure 7.5.3: Plot of $\frac{dP}{dt}$ vs. P for $\frac{dP}{dt} = \frac{1}{100}(1.5 \cdot 10^5 - P)$.

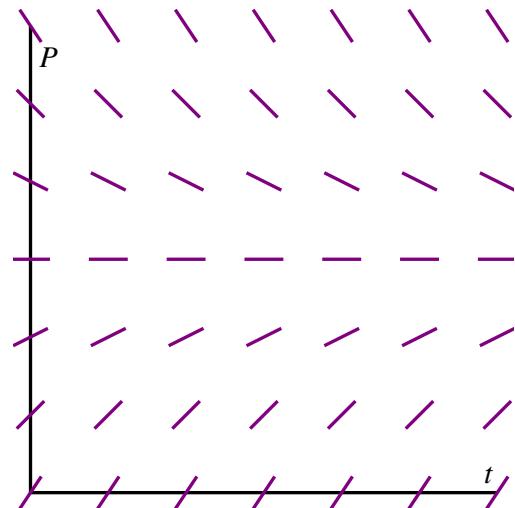


Figure 7.5.4: Plot of the slope field for $\frac{dP}{dt} = \frac{1}{100}(1.5 \cdot 10^5 - P)$.

These plots both show that $P = 1.5 \cdot 10^5$ is a stable equilibrium. Therefore, we should expect that the amount of pollution in Lake Michigan will stabilize near $1.5 \cdot 10^5$ cubic meters of pollution.

Next, assuming that there is initially no pollution in the lake, we will solve the initial value problem

$$\frac{dP}{dt} = \frac{1}{100}(1.5 \cdot 10^5 - P), \quad P(0) = 0.$$

Separating variables, we find that

$$\frac{1}{1.5 \cdot 10^5 - P} \frac{dP}{dt} = \frac{1}{100}.$$

Integrating with respect to t , we have

$$\int \frac{1}{1.5 \cdot 10^5 - P} \frac{dP}{dt} dt = \int \frac{1}{100} dt,$$

and thus changing variables on the left and antidifferentiating on both sides, we find that

$$\int \frac{dP}{1.5 \cdot 10^5 - P} = \int \frac{1}{100} dt$$

$$-\ln|1.5 \cdot 10^5 - P| = \frac{1}{100}t + C$$

Finally, multiplying both sides by -1 and using the definition of the logarithm, we find that

$$1.5 \cdot 10^5 - P = Ce^{-t/100}. \quad (7.5.1)$$

This is a good time to determine the constant C . Since $P = 0$ when $t = 0$, we have

$$1.5 \cdot 10^5 - 0 = Ce^0 = C,$$

so $C = 1.5 \cdot 10^5$.

Using this value of C in Equation (7.5.1) and solving for P , we arrive at the solution

$$P(t) = 1.5 \cdot 10^5(1 - e^{-t/100}).$$

Superimposing the graph of P on the slope field we saw in Figure 7.5.4, we see, as shown in Figure 7.5.5

We see that, as expected, the amount of plastic pollution stabilizes around $1.5 \cdot 10^5$ cubic meters.

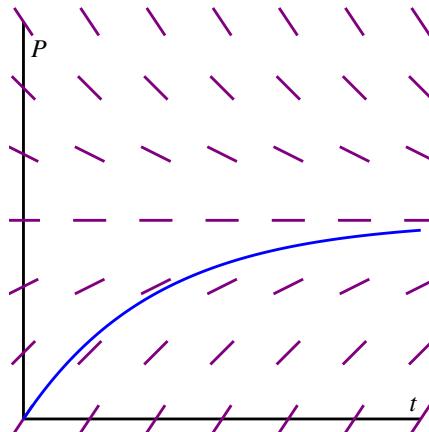


Figure 7.5.5: The solution $P(t)$ and the slope field for the differential equation $\frac{dP}{dt} = \frac{1}{100}(1.5 \cdot 10^5 - P)$.

There are many important lessons to learn from Example 7.5.2. Foremost is how we can develop a differential equation by thinking about the “total rate = rate in - rate out” model. In addition, we note how we can bring together all of our available understanding (plotting $\frac{dP}{dt}$ vs. P , creating a slope field, solving the differential equation) to see how the differential equation describes the behavior of a changing quantity.

We can also explore what happens when certain aspects of the problem change. For instance, let’s suppose we are at a time when the plastic pollution entering Lake Michigan has stabilized at $1.5 \cdot 10^5$ cubic meters, and that new legislation is passed to prevent this type of

pollution entering the lake. So, there is no longer any inflow of plastic pollution to the lake. How does the amount of plastic pollution in Lake Michigan now change? For example, how long does it take for the amount of plastic pollution in the lake to halve?

Resetting $t = 0$ at this time, we now have the initial value problem

$$\frac{dP}{dt} = -\frac{1}{100}P, \quad P(0) = 1.5 \cdot 10^5.$$

It is a straightforward and familiar exercise to find that the solution to this equation is $P(t) = 1.5 \cdot 10^5 e^{-t/100}$. The time that it takes for half of the pollution to flow out of the lake is given by T where $P(T) = 0.75 \cdot 10^5$. Thus, we must solve the equation

$$0.75 \cdot 10^5 = 1.5 \cdot 10^5 e^{-T/100},$$

or

$$\frac{1}{2} = e^{-T/100}.$$

It follows that

$$T = -100 \ln\left(\frac{1}{2}\right) \approx 69.3 \text{ years.}$$

In the upcoming activities, we explore some other natural settings in which differential equations model changing quantities.

Example 7.5.6. Suppose you have a bank account that grows by 5% every year. Let $A(t)$ be the amount of money in the account in year t .

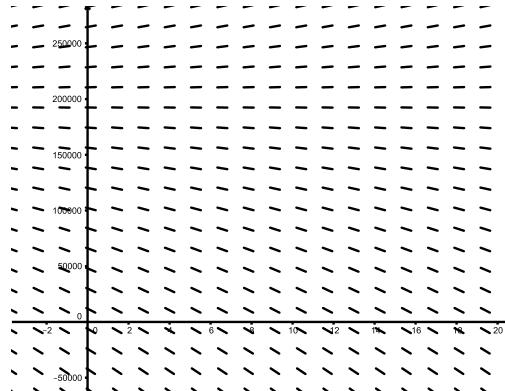
- What is the rate of change of A with respect to t ?
- Suppose that you are also withdrawing \$10,000 per year. Write a differential equation that expresses the total rate of change of A .
- Sketch a slope field for this differential equation, find any equilibrium solutions, and identify them as either stable or unstable. Write a sentence or two that describes the significance of the stability of the equilibrium solution.
- Suppose that you initially deposit \$100,000 into the account. How long does it take for you to deplete the account?
- What is the smallest amount of money you would need to have in the account to guarantee that you never deplete the money in the account?
- If your initial deposit is \$300,000, how much could you withdraw every year without depleting the account?

Hint.

- Small hints for each of the prompts above.

Answer.

- a. $\frac{dA}{dt} = 0.05A$.
- b. $\frac{dA}{dt} = 0.05A - 10000$.
- c.



The only equilibrium solution is $A = 200000$.

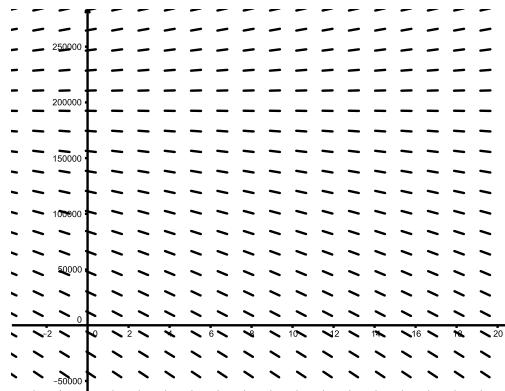
- d. $t = 20 \ln(2) \approx 13.86$ years.
- e. At least \$200000.
- f. Up to \$15000 every year.

Solution. Let $A(t)$ be the amount of money in a bank account in year t that grows by 5% every year.

- a. The rate of change of A with respect to t is $\frac{dA}{dt} = 0.05A$.
- b. If \$10000 is being withdrawn every year, then

$$\frac{dA}{dt} = 0.05A - 10000.$$

- c. Following is a slope field for the differential equation in (b).



Equilibrium solutions occur where $\frac{dA}{dt} = 0$ and so the only equilibrium solution is

$A = 200000$. This means that if you have \$200000 in the account, then the amount of money in the account will remain constant at \$200000.

- d. Suppose we initially have \$100000 in the account. We will solve the intitial value problem

$$\frac{dA}{dt} = 0.05A - 10000, A(0) = 100000.$$

Using separation of variables, we find that

$$A = 200000 + Ce^{t/20},$$

and since $A(0) = 100000$, we have

$$A = 200000 - 100000e^{t/20}.$$

To determine when the account will be depleted, we solve the equation $A = 0$ for t . Since $200000 - 100000e^{t/20} = 0$, it follows $e^{t/20} = 2$ and thus $t = 20 \ln(2)$, so the account will be depleted in about 13.86 years.

- e. You must have at least \$200000 in the account to make sure the account will never be depleted.
- f. Now suppose the intitial deposit is \$300000 and we want to what amount w we can withdraw every year without depleting the account. So the initial value problem is

$$\frac{dA}{dt} = 0.05A - w, A(0) = 300000.$$

The solution to this initial value problem is

$$A = 20w + (300000 - 20w)e^{t/20}.$$

To make sure the account is never depleted, we need $(300000 - 20w)$ to be greater than or equal to 0. Solving the inequality $300000 - 2w \geq 0$, we obtain $w \leq 15000$. We can withdraw up to \$15000 every year and the account will not be depleted.

Example 7.5.7. A dose of morphine is absorbed from the bloodstream of a patient at a rate proportional to the amount in the bloodstream.

- Write a differential equation for $M(t)$, the amount of morphine in the patient's bloodstream, using k as the constant proportionality.
- Assuming that the initial dose of morphine is M_0 , solve the initial value problem to find $M(t)$. Use the fact that the half-life for the absorption of morphine is two hours to find the constant k .
- Suppose that a patient is given morphine intravenously at the rate of 3 milligrams per hour. Write a differential equation that combines the intravenous administration of morphine with the body's natural absorption.
- Find any equilibrium solutions and determine their stability.
- Assuming that there is initially no morphine in the patient's bloodstream, solve the initial value problem to determine $M(t)$. What happens to $M(t)$ after a very long time?

- f. To what rate should a doctor reduce the intravenous rate so that there is eventually 7 milligrams of morphine in the patient's bloodstream?

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a. $\frac{dM}{dt} = -kM$, where k is a positive constant.
 b. $k = -\frac{1}{2} \ln\left(\frac{1}{2}\right) \approx 0.34657$.
 c. $\frac{dM}{dt} = 3 - kM$, where k is a positive constant.
 d. The equilibrium solution $mM = \frac{3}{k}$ is stable.
 e. $M = \frac{3}{k} \left(1 - e^{-kt}\right)$.
 f. About 2.426 milligrams per hour.

Solution.

- a. We will use the differential equation $\frac{dM}{dt} = -kM$, where k is a positive constant.
 b. With an initial condition of $M = M_0$ when $t = 0$, the solution for the initial value problem is

$$M = M_0 e^{-kt}.$$

Using a half-life of 2 hours, we see that $M = \frac{1}{2}M_0$ when $t = 2$. This gives $\frac{1}{2}M_0 = M_0 e^{-2k}$, from which it follows that $k = -\frac{1}{2} \ln\left(\frac{1}{2}\right) \approx 0.34657$.

- c. Noting that the rate of change of the drug in the body will be the difference between the positive contributions of the dosage of 3 milligrams per hour and the decay of kM per hour, it follows that $\frac{dM}{dt} = 3 - kM$, where k is a positive constant.
 d. To find the equilibrium solutions, we solve the equation $\frac{dM}{dt} = 0$ for M . This gives $M = \frac{3}{k}$. The graph of $\frac{dM}{dt}$ as a function of M is a straight line with a negative slope. (The slope is $-k$.) So when $M < \frac{3}{k}$, $\frac{dM}{dt} > 0$ and the M is an increasing function of t . When $M > \frac{3}{k}$, $\frac{dM}{dt} < 0$ and the M is a decreasing function of t . This means that the equilibrium solution $mM = \frac{3}{k}$ is a stable equilibrium solution.
 e. We use the initial condition $M = 0$ when $t = 0$ and solve the corresponding IVP. Separating variables,

$$\frac{1}{3 - kM} \frac{dM}{dt} = 1,$$

so

$$\int \frac{1}{3 - kM} dM = \int 1 dt,$$

and thus

$$-\frac{1}{k} \ln |3 - kM| = t + c.$$

Solving for M in the usual way gives us

$$M = \frac{1}{k} (3 - Ae^{-kt}),$$

where A is a constant that arises from the integration constant. Using the initial condition, we obtain $0 = \frac{1}{k}(3 - A)$, so $A = 3$. Hence, the solution to the initial condition problem is $M = \frac{1}{k} (3 - 3e^{-kt}) = \frac{3}{k} (1 - e^{-kt})$.

- f. We let q be the unknown intravenous rate of morphine (in milligrams per hour). Notice that if we replace 3 with q in the differential equation and obtain $\frac{dM}{dt} = q - kM$, the solution process is the same and the equilibrium solution is $M = \frac{q}{k}$. So if we want the equilibrium solution to be $M = 7$, we then have $\frac{q}{k} = 7$, so $q = 7k \approx 2.426$. The intravenous rate of morphine should be 2.426 milligrams per hour.

7.5.2 Summary

- Differential equations arise in a situation when we understand how various factors cause a quantity to change.
- We may use the tools we have developed so far—slope fields, Euler’s methods, and our method for solving separable equations—to understand a quantity described by a differential equation.

7.5.3 Exercises

1. (Mixing problem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Mixing problem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Population growth problem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Radioactive decay problem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Investment problem) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. Congratulations, you just won the lottery! In one option presented to you, you will be paid one million dollars a year for the next 25 years. You can deposit this money in an account that will earn 5% each year.
 - a. Set up a differential equation that describes the rate of change in the amount of money in the account. Two factors cause the amount to grow—first, you are depositing one million dollars per year and second, you are earning 5% interest.

- b. If there is no amount of money in the account when you open it, how much money will you have in the account after 25 years?
- c. The second option presented to you is to take a lump sum of 10 million dollars, which you will deposit into a similar account. How much money will you have in that account after 25 years?
- d. Do you prefer the first or second option? Explain your thinking.
- e. At what time does the amount of money in the account under the first option overtake the amount of money in the account under the second option?

Answer.

- a. $\frac{dA}{dt} = 1 + 0.05A$
- b. $A(25) = 49.80686$ million dollars.
- c. $A(25) = 34.90343$ million dollars.
- d. The first.
- e. $t = 20 \ln(2) \approx 13.86$ years.

Solution.

- a. Let $A(t)$ be the amount of money (in millions) in the account at time t . We know that the rate of change of A is affected by the addition of 1 million dollars a year, plus the interest earned, which is $0.05A$. Thus,

$$\frac{dA}{dt} = 1 + 0.05A$$

- b. If there is no amount of money in the account when you open it, that means $A(0) = 0$. Solving the initial value problem, we can find how much money will be in the account after 25 years. Note that $1 + 0.05A = 0.05(20 + A)$, so we have

$$\frac{1}{20 + A} \frac{dA}{dt} = 0.05$$

and thus

$$\int \frac{1}{20 + A} dA = \int 0.05 dt.$$

Integrating (and noting that $A(t) > 0$ for all t),

$$\ln(20 + A) = 0.05t + C,$$

so solving for A in the standard way,

$$20 + A = Ke^{0.05t}$$

and thus $A(t) = Ke^{0.05t} - 20$. Since $A(0) = 0$, $K = 20$, and we've shown

$$A(t) = 20(e^{0.05t} - 1).$$

It follows that $A(25) = 49.80686$ million dollars.

- c. For this option, we know that $\frac{dA}{dt} = 0.05A$, $A(0) = 10$, and thus by the standard solution to this IVP,

$$A(t) = 10e^{0.05t}$$

from which we see that $A(25) = 34.90343$ million dollars.

- d. The first option is obviously better because it leads to more money in the long term.
e. The amount of money in the account under the first option overtakes the amount of money in the account under the second option when $20(e^{0.05t} - 1) = 10e^{0.05t}$. Solving for t , we find that $10e^{0.05t} = 20$, so $e^{0.05t} = 2$, and thus

$$t = 20 \ln(2) \approx 13.86 \text{ years.}$$

7. When a skydiver jumps from a plane, gravity causes her downward velocity to increase at the rate of $g \approx 9.8$ meters per second squared. At the same time, wind resistance causes her velocity to decrease at a rate proportional to the velocity.

- a. Using k to represent the constant of proportionality, write a differential equation that describes the rate of change of the skydiver's velocity.
b. Find any equilibrium solutions and decide whether they are stable or unstable. Your result should depend on k .
c. Suppose that the initial velocity is zero. Find the velocity $v(t)$.
d. A typical terminal velocity for a skydiver falling face down is 54 meters per second. What is the value of k for this skydiver?
e. How long does it take to reach 50% of the terminal velocity?

Answer.

- a. $\frac{dv}{dt} = 9.8 - kv$
b. $v = \frac{9.8}{k}$ is a stable equilibrium.
c. $v(t) = \frac{9.8 - 9.8e^{-kt}}{k}$
d. $k = 9.8/54 \approx 0.181481$.
e. $t = \frac{\ln(0.5)}{-0.181481} \approx 3.1894$ seconds.

Solution.

- a. Let the skydiver's downward velocity at time t be given by $v(t)$. Since gravity causes her downward velocity to increase at the rate of $g \approx 9.8$, there is a factor of $+9.8$ meters per second squared in the differential equation for $\frac{dv}{dt}$, while the factor of wind resistance causing the velocity to decrease contributes the negative factor $-kv$. Thus, the differential equation is

$$\frac{dv}{dt} = 9.8 - kv$$

- b. To find the equilibrium solutions, we set $\frac{dv}{dt} = 0$ and solve for v . Doing so, $0 = 9.8 - kv$, so $v = \frac{9.8}{k}$. Moreover, we note that for $v < \frac{9.8}{k}$, $\frac{dv}{dt} > 0$, while for $\frac{dv}{dt} > \frac{9.8}{k}$, $\frac{dv}{dt} < 0$. Thus, for velocities just above the equilibrium, the velocity will decrease and for velocities just below the equilibrium, the velocity will increase, making $v = \frac{9.8}{k}$ a stable equilibrium.

- c. We assume that $v(0) = 0$ and solve the IVP with the differential equation we developed. Taking the differential equation

$$\frac{dv}{dt} = 9.8 - kv$$

and separating the variables, we have

$$\frac{1}{9.8 - kv} \frac{dv}{dt} = 1.$$

Integrating with respect to t ,

$$\int \frac{1}{9.8 - kv} dv = \int 1 dt,$$

and using a standard substitution, we have

$$-\frac{1}{k} \ln |9.8 - kv| = t + C.$$

Solving for v , we have $\ln |9.8 - kv| = -kt + C_1$ (where $C_1 = -Ck$, or

$$|9.8 - kv| = e^{-kt+C_1} = C_2 e^{-kt}.$$

Incorporating the absolute value into the constant C_2 in the usual way, we have for some constant A that

$$9.8 - kv = Ae^{-kt}$$

and thus

$$v = \frac{9.8 - Ae^{-kt}}{k}.$$

Setting $v(0) = 0$, we can also determine A , since then

$$0 = \frac{9.8 - A}{k}$$

so $A = 9.8$. The solution to the IVP is therefore

$$v(t) = \frac{9.8 - 9.8e^{-kt}}{k}$$

- d. If the skydiver's terminal velocity is 54 meters per second, this means that

$$\lim_{t \rightarrow \infty} v(t) = 54$$

Using the solution we found in (c), we observe that

$$\lim_{t \rightarrow \infty} \frac{9.8 - 9.8e^{-kt}}{k} = \frac{9.8}{k}$$

since $e^{-kt} \rightarrow 0$ as $t \rightarrow \infty$. Thus, $54 = \frac{9.8}{k}$, so $k = 9.8/54 \approx 0.181481$.

- e. To find how long does it take to reach 50% of the terminal velocity, we set $v(t) = \frac{9.8 - 9.8e^{-0.181481t}}{0.181481} = 0.5(54)$ and solve for t . We have

$$9.8 - 9.8e^{-0.181481t} = 0.5(54)(0.181481),$$

so

$$9.8 - 9.8e^{-0.181481t} = 4.9,$$

and thus $9.8e^{-0.181481t} = 4.9$, which tells us

$$e^{-0.181481t} = 0.5.$$

Solving for t , $t = \frac{\ln(0.5)}{-0.181481} \approx 3.1894$ seconds.

8. During the first few years of life, the rate at which a baby gains weight is proportional to the reciprocal of its weight.

- Express this fact as a differential equation.
- Suppose that a baby weighs 8 pounds at birth and 9 pounds one month later. How much will he weigh at one year?
- Do you think this is a realistic model for a long time?

Answer.

- $\frac{dw}{dt} = \frac{k}{w}$.
- $w(t) = \sqrt{7t + 64}$; $w(12) = \sqrt{148} \approx 12.17$ pounds.
- The model is unrealistic.

Solution.

- a. Since the baby's weight, $w(t)$ (in pounds at month t), grows at a rate proportional to its reciprocal, we have

$$\frac{dw}{dt} = \frac{k}{w}.$$

- b. Given that $W(0) = 8$ and $W(1) = 9$, we can both solve the IVP and determine k . To start, we separate variables and solve the differential equation by writing

$$w \frac{dw}{dt} = k$$

and integrating to get

$$\int w \frac{dw}{dt} dt = \int k dt.$$

It follows that

$$\frac{1}{2}w^2 = kt + C$$

and therefore $w(t) = \sqrt{2kt + C_1}$ (where $C_1 = 2C$ and we take the positive square root because we know $w(t) > 0$). Applying the initial condition $w(0) = 0$, we find that $8 = \sqrt{C_1}$, so $C_1 = 64$ and

$$w(t) = \sqrt{2kt + 64}.$$

To find k , we use the additional information that $w(1) = 9$, so

$$9 = \sqrt{2k + 64}$$

which implies that $81 = 2k + 64$ and thus $k = 17/2$. Thus,

$$w(t) = \sqrt{7t + 64}.$$

At one year, according to the model the baby weighs $w(12) = \sqrt{148} \approx 12.17$ pounds.

- c. While the function $w(t) = \sqrt{7t + 64}$ grows without bound as t increases, the model is unrealistic more because the function initially doesn't grow fast enough. For example, the model predicts that at age 10, $w(120) \approx 30.06$ pounds, and at age 20, $w(120) = 41.76$ pounds. A realistic model would need to grow much more quickly early in life and then level off around age 20.
- 9. Suppose that you have a water tank that holds 100 gallons of water. A briny solution, which contains 20 grams of salt per gallon, enters the tank at the rate of 3 gallons per minute. At the same time, the solution is well mixed, and water is pumped out of the tank at the rate of 3 gallons per minute.
 - a. Since 3 gallons enters the tank every minute and 3 gallons leaves every minute, what can you conclude about the volume of water in the tank?
 - b. How many grams of salt enters the tank every minute?
 - c. Suppose that $S(t)$ denotes the number of grams of salt in the tank in minute t . How many grams are there in each gallon in minute t ?
 - d. Since water leaves the tank at 3 gallons per minute, how many grams of salt leave the tank each minute?
 - e. Write a differential equation that expresses the total rate of change of S .
 - f. Identify any equilibrium solutions and determine whether they are stable or unstable.
 - g. Suppose that there is initially no salt in the tank. Find the amount of salt $S(t)$ in minute t .
 - h. What happens to $S(t)$ after a very long time? Explain how you could have predicted this only knowing how much salt there is in each gallon of the briny solution that enters the tank.

Answer.

- a. The inflow and outflow are at the same rate.
- b. 60 grams per minute.
- c. $\frac{S(t) \text{ grams}}{100 \text{ gallon}}$
- d. $\frac{3S(t) \text{ grams}}{100 \text{ minute}}$.
- e. $\frac{dS}{dt} = 60 - \frac{3}{100}S$.
- f. $S = 2000$ is a stable equilibrium solution.
- g. $S(t) = 2000 - 2000e^{-\frac{3}{100}t}$.
- h. $S(t) \rightarrow 2000$.

Solution.

- The volume of solution in the tank remains constant because the inflow and outflow are at the same rate.
- Salt enters the tank with the inflow, which occurs at a rate of 3 gallons per minute at a concentration of 20 grams per gallon. Observing that

$$20 \frac{\text{grams}}{\text{gallon}} \cdot 3 \frac{\text{gallons}}{\text{minute}} = 60 \frac{\text{grams}}{\text{minute}},$$

we see that salt is entering the tank at a rate of 60 grams per minute.

- Letting $S(t)$ denotes the number of grams of salt in the tank at time t , since there are 100 gallons of solution in the tank at any time t , the concentration of salt in the tank in grams per gallon is

$$\frac{S(t)}{100} \frac{\text{grams}}{\text{gallon}}$$

at any time t .

- With water leaving the tank at 3 gallons per minute, it follows that salt is leaving the tank at a rate of

$$\frac{S(t)}{100} \frac{\text{grams}}{\text{gallon}} \cdot 3 \frac{\text{gallons}}{\text{minute}} = \frac{3S(t)}{100} \frac{\text{grams}}{\text{minute}}.$$

- Taking into account the rates at which salt enters and leaves the tank as established in (b) and (d) and noting that the units on $\frac{dS}{dt}$ are grams per minute, we have shown that

$$\frac{dS}{dt} = 60 - \frac{3}{100}S.$$

- Setting $60 - \frac{3}{100}S = 0$, we find that $S = 2000$ is the only equilibrium solution. This makes sense because the inflow is carrying 20 grams of salt per gallon and thus when the tank reaches this concentration, there will be $20 \cdot 100$ total grams of salt. This equilibrium is stable (both intuitively and because $\frac{dS}{dt}$ is positive for $S < 2000$ and negative for $S > 2000$).

- Assuming that $S(0) = 0$, we can solve the IVP with $\frac{dS}{dt} = 60 - \frac{3}{100}S$. For somewhat simpler computations, we write $60 - \frac{3}{100}S = -\frac{3}{100}(S - 2000)$. Separating variables,

$$\frac{1}{S - 2000} \frac{dS}{dt} = -\frac{3}{100}$$

and integrating with respect to t ,

$$\int \frac{1}{S - 2000} dS = - \int \frac{3}{100} dt$$

so that

$$\ln |S - 2000| = -\frac{3}{100}t + C.$$

Solving for S in the usual way,

$$S = 2000 + Ke^{-\frac{3}{100}t}$$

and using the initial condition that $S(0) = 0$, we find $K = -2000$ so that

$$S(t) = 2000 - 2000e^{-\frac{3}{100}t}.$$

- h. As $t \rightarrow \infty$, we know $e^{-\frac{3}{100}t} \rightarrow 0$, and thus $S(t) \rightarrow 2000$. This makes sense because the inflow carries 20 grams of salt per liter continuously. As time goes on, the concentration of solution in the tank should tend to that same concentration, and with 100 gallons of solution, the amount of salt will tend to $20 \cdot 100 = 2000$ grams. This value also matches the stable equilibrium that we found in (f).

7.6 Population Growth and the Logistic Equation

Motivating Questions

- How can we use differential equations to realistically model the growth of a population?
 - How can we assess the accuracy of our models?
-

The growth of the earth's population is one of the pressing issues of our time. Will the population continue to grow? Or will it perhaps level off at some point, and if so, when? In this section, we look at two ways in which we may use differential equations to help us address these questions.

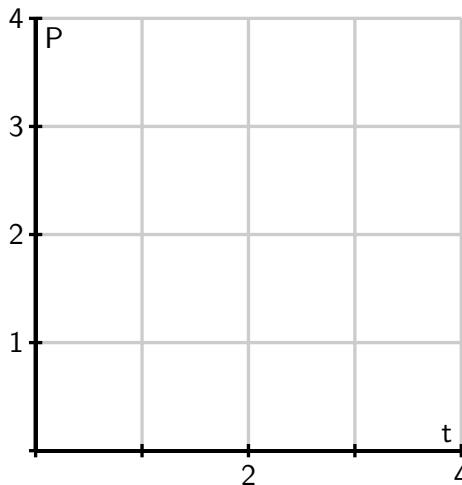
Before we begin, let's consider again two important differential equations that we have seen in earlier work this chapter.

Example 7.6.1. Recall that one model for population growth states that a population grows at a rate proportional to its size.

- We begin with the differential equation

$$\frac{dP}{dt} = \frac{1}{2}P.$$

Sketch a slope field below as well as a few typical solutions on the axes provided.

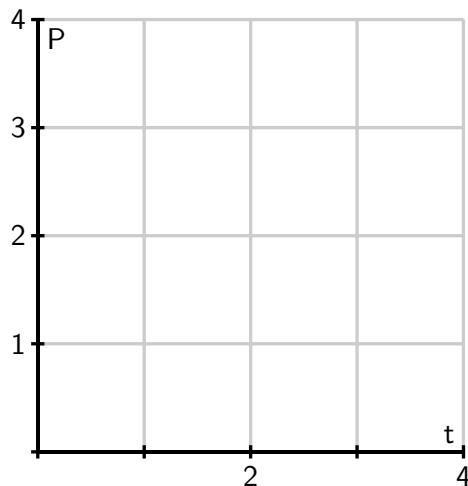


- Find all equilibrium solutions of the equation $\frac{dP}{dt} = \frac{1}{2}P$ and classify them as stable or unstable.
- If $P(0)$ is positive, describe the long-term behavior of the solution to $\frac{dP}{dt} = \frac{1}{2}P$.

d. Let's now consider a modified differential equation given by

$$\frac{dP}{dt} = \frac{1}{2}P(3 - P).$$

As before, sketch a slope field as well as a few typical solutions on the following axes provided.



e. Find any equilibrium solutions and classify them as stable or unstable.

f. If $P(0)$ is positive, describe the long-term behavior of the solution.

7.6.1 The earth's population

We will now begin studying the earth's population. To get started, in Table 7.6.2 are some data for the earth's population in recent years that we will use in our investigations.

Year	1998	1999	2000	2001	2002	2005	2006	2007	2008	2009	2010
Pop (billions)	5.932	6.008	6.084	6.159	6.234	6.456	6.531	6.606	6.681	6.756	6.831

Table 7.6.2: Some recent population data for planet Earth.

Example 7.6.3. Our first model will be based on the following assumption:

The rate of change of the population is proportional to the population.

On the face of it, this seems pretty reasonable. When there is a relatively small number of people, there will be fewer births and deaths so the rate of change will be small. When there

is a larger number of people, there will be more births and deaths so we expect a larger rate of change.

If $P(t)$ is the population t years after the year 2000, we may express this assumption as

$$\frac{dP}{dt} = kP$$

where k is a constant of proportionality.

- Use the data in the table to estimate the derivative $P'(0)$ using a central difference. Assume that $t = 0$ corresponds to the year 2000.
- What is the population $P(0)$?
- Use your results from (a) and (b) to estimate the constant of proportionality k in the differential equation.
- Now that we know the value of k , we have the initial value problem

$$\frac{dP}{dt} = kP, P(0) = 6.084.$$

Find the solution to this initial value problem.

- What does your solution predict for the population in the year 2010? Is this close to the actual population given in the table?
- When does your solution predict that the population will reach 12 billion?
- What does your solution predict for the population in the year 2500?
- Do you think this is a reasonable model for the earth's population? Why or why not? Explain your thinking using a couple of complete sentences.

Hint.

- Small hints for each of the prompts above.

Answer.

- $P'(0) \approx 0.0755$.
- $P(0) = 6.084$.
- $k \approx 0.012041$.
- $P(t) = 6.084e^{0.012041t}$.
- $P(10) \approx 6.8878$.
- $t = \frac{1}{0.012041} \ln\left(\frac{12}{6.084}\right) \approx 56.41$, or in the year 2056.
- $P(500) \approx 3012.3$ billion.

Solution. We let $P(t)$ be the population after year 2000 with $\frac{dP}{dt} = kP$, where k is a constant of proportionality.

a. Using the data in the table,

$$P'(0) \approx \frac{P(1) - P(-1)}{2} \approx \frac{6.159 - 6.008}{2} \approx 0.0755.$$

- b. Since $t = 0$ corresponds to the year 2000, $P(0) = 6.084$.
- c. Using $\frac{dP}{dt} = kP$ and the preceding values at $t = 0$, we have $0.0755 \approx k(6.084)$, so $k \approx 0.012041$.
- d. The solution for the initial value problem is $P(t) = 6.084e^{0.012041t}$.
- e. The year 2010 corresponds to $t = 10$. So $P(10) \approx 6.8878$. The model predicts that the population in 2010 will be about 6.888 billion.
- f. The population will be 12 billion when $P(t) = 12$. So we solve the equation $6.084e^{0.012041t} = 12$ for t . We see that $e^{0.012041t} = \frac{12}{6.084}$, so $t = \frac{1}{0.012041} \ln\left(\frac{12}{6.084}\right) \approx 56.41$. The population will reach 12 billion during the year 2056.
- g. The year 2500 corresponds to $t = 500$. So $P(500) \approx 3012.3$. The model predicts that the population in 2500 will be about 3012 billion or about 3 012 300 000 000. This is rather unreasonably large.

Our work in Example 7.6.3 shows that the exponential model is fairly accurate for years relatively close to 2000. However, if we go too far into the future, the model predicts increasingly large rates of change, which causes the population to grow arbitrarily large. This does not make much sense since it is unrealistic to expect that the earth would be able to support such a large population.

The constant k in the differential equation has an important interpretation. Let's rewrite the differential equation $\frac{dP}{dt} = kP$ by solving for k , so that we have

$$k = \frac{dP/dt}{P}.$$

We see that k is the ratio of the rate of change to the population; in other words, it is the contribution to the rate of change from a single person. We call this the *per capita growth rate*.

In the exponential model we introduced in Example 7.6.3, the per capita growth rate is constant. This means that when the population is large, the per capita growth rate is the same as when the population is small. It is natural to think that the per capita growth rate should decrease when the population becomes large, since there will not be enough resources to support so many people. We expect it would be a more realistic model to assume that the per capita growth rate depends on the population P .

In the previous example, we computed the per capita growth rate in a single year by computing k , the quotient of $\frac{dP}{dt}$ and P (which we did for $t = 0$). If we return to the data in Table 7.6.2 and compute the per capita growth rate over a range of years, we generate the

data shown in Figure 7.6.4, which shows how the per capita growth rate is a function of the population, P .

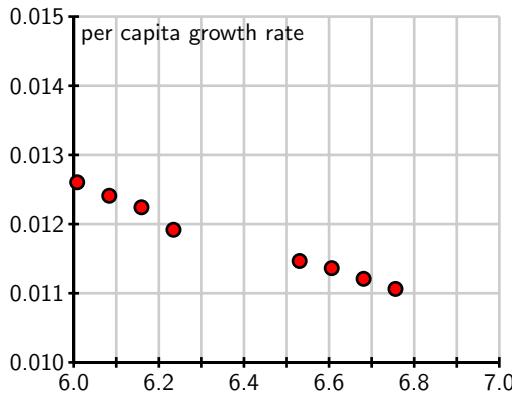


Figure 7.6.4: A plot of per capita growth rate vs. population P .

From the data, we see that the per capita growth rate appears to decrease as the population increases. In fact, the points seem to lie very close to a line, which is shown at two different scales in Figure 7.6.5.

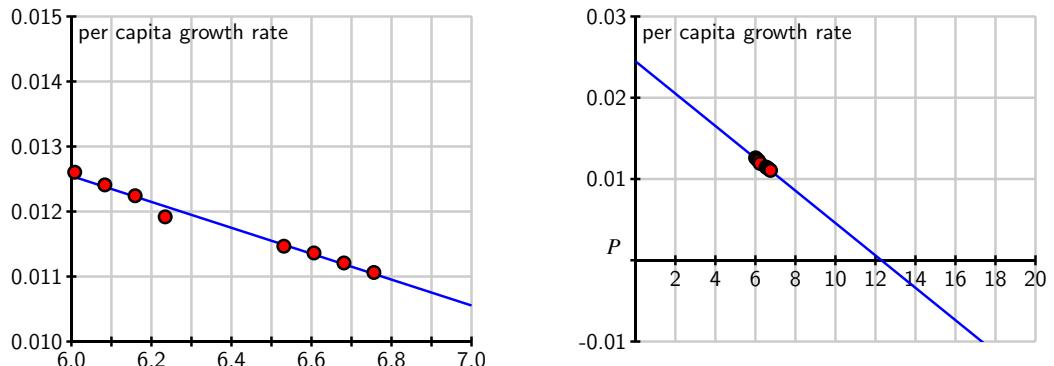


Figure 7.6.5: The line that approximates per capita growth as a function of population, P .

Looking at this line carefully, we can find its equation to be

$$\frac{dP/dt}{P} = 0.025 - 0.002P.$$

If we multiply both sides by P , we arrive at the differential equation

$$\frac{dP}{dt} = P(0.025 - 0.002P).$$

Graphing the dependence of dP/dt on the population P , we see that this differential equation demonstrates a quadratic relationship between $\frac{dP}{dt}$ and P , as shown in Figure 7.6.6.

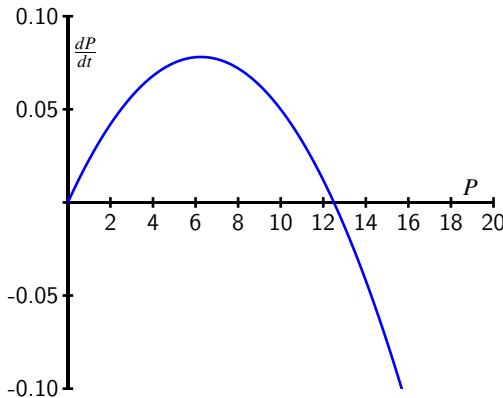


Figure 7.6.6: A plot of $\frac{dP}{dt}$ vs. P for the differential equation $\frac{dP}{dt} = P(0.025 - 0.002P)$.

The equation $\frac{dP}{dt} = P(0.025 - 0.002P)$ is an example of the *logistic equation*, and is the second model for population growth that we will consider. We expect that it will be more realistic, because the per capita growth rate is a decreasing function of the population.

Indeed, the graph in Figure 7.6.6 shows that there are two equilibrium solutions, $P = 0$, which is unstable, and $P = 12.5$, which is a stable equilibrium. The graph shows that any solution with $P(0) > 0$ will eventually stabilize around 12.5. Thus, our model predicts the world's population will eventually stabilize around 12.5 billion.

A prediction for the long-term behavior of the population is a valuable conclusion to draw from our differential equation. We would, however, also like to answer some quantitative questions. For instance, how long will it take to reach a population of 10 billion? To answer this question, we need to find an explicit solution of the equation.

7.6.2 Solving the logistic differential equation

Since we would like to apply the logistic model in more general situations, we state the logistic equation in its more general form,

$$\frac{dP}{dt} = kP(N - P). \quad (7.6.1)$$

The equilibrium solutions here are $P = 0$ and $1 - \frac{P}{N} = 0$, which shows that $P = N$. The equilibrium at $P = N$ is called the *carrying capacity* of the population for it represents the stable population that can be sustained by the environment.

We now solve the logistic equation (7.6.1). The equation is separable, so we separate the variables

$$\frac{1}{P(N - P)} \frac{dP}{dt} = k,$$

and integrate to find that

$$\int \frac{1}{P(N - P)} dP = \int k dt.$$

To find the antiderivative on the left, we use the partial fraction decomposition

$$\frac{1}{P(N - P)} = \frac{1}{N} \left[\frac{1}{P} + \frac{1}{N - P} \right].$$

Now we are ready to integrate, with

$$\int \frac{1}{N} \left[\frac{1}{P} + \frac{1}{N - P} \right] dP = \int k dt.$$

On the left, observe that N is constant, so we can remove a factor of $\frac{1}{N}$ and antidifferentiate to find that

$$\frac{1}{N} (\ln |P| - \ln |N - P|) = kt + C.$$

Multiplying both sides of this last equation by N and using a rule of logarithms, we next find that

$$\ln \left| \frac{P}{N - P} \right| = kNt + C.$$

From the definition of the logarithm, replacing e^C with C , and letting C absorb the absolute value signs, we now know that

$$\frac{P}{N - P} = Ce^{kNt}.$$

At this point, all that remains is to determine C and solve algebraically for P .

If the initial population is $P(0) = P_0$, then it follows that $C = \frac{P_0}{N - P_0}$, so

$$\frac{P}{N - P} = \frac{P_0}{N - P_0} e^{kNt}.$$

We will solve this equation for P by multiplying both sides by $(N - P)(N - P_0)$ to obtain

$$\begin{aligned} P(N - P_0) &= P_0(N - P)e^{kNt} \\ &= P_0Ne^{kNt} - P_0Pe^{kNt}. \end{aligned}$$

Solving for P_0Ne^{kNt} , expanding, and factoring, it follows that

$$\begin{aligned} P_0Ne^{kNt} &= P(N - P_0) + P_0Pe^{kNt} \\ &= P(N - P_0 + P_0e^{kNt}). \end{aligned}$$

Dividing to solve for P , we see that

$$P = \frac{P_0Ne^{kNt}}{N - P_0 + P_0e^{kNt}}.$$

Finally, we choose to multiply the numerator and denominator by $\frac{1}{P_0}e^{-kNt}$ to obtain

$$P(t) = \frac{N}{\left(\frac{N - P_0}{P_0} \right) e^{-kNt} + 1}.$$

While that was a lot of algebra, notice the result: we have found an explicit solution to the logistic equation.

Solution to the Logistic Equation

The solution to the initial value problem

$$\frac{dP}{dt} = kP(N - P), \quad P(0) = P_0,$$

is

$$P(t) = \frac{N}{\left(\frac{N-P_0}{P_0}\right) e^{-kNt} + 1}. \quad (7.6.2)$$

For the logistic equation describing the earth's population that we worked with earlier in this section, we have

$$k = 0.002, N = 12.5, \text{ and } P_0 = 6.084.$$

This gives the solution

$$P(t) = \frac{12.5}{1.0546e^{-0.025t} + 1},$$

whose graph is shown in Figure 7.6.7.

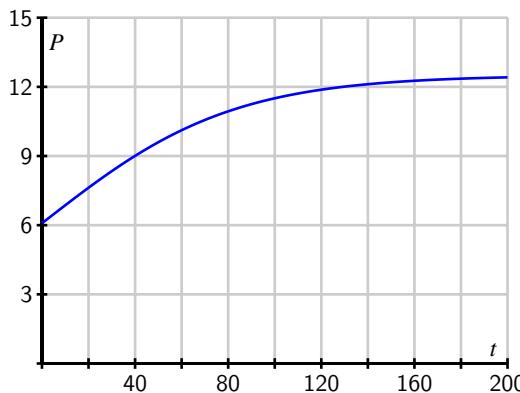


Figure 7.6.7: The solution to the logistic equation modeling the earth's population.

The graph shows the population leveling off at 12.5 billion, as we expected, and that the population will be around 10 billion in the year 2050. These results, which we have found using a relatively simple mathematical model, agree fairly well with predictions made using a much more sophisticated model developed by the United Nations.

The logistic equation is good for modeling any situation in which limited growth is possible. For instance, it could model the spread of a flu virus through a population contained on a cruise ship, the rate at which a rumor spreads within a small town, or the behavior of an animal population on an island. Through our work in this section, we have completely solved the logistic equation, regardless of the values of the constants N , k , and P_0 . Anytime we encounter a logistic equation, we can apply the formula we found in Equation (7.6.2).

Example 7.6.8. Consider the logistic equation

$$\frac{dP}{dt} = kP(N - P)$$

with the graph of $\frac{dP}{dt}$ vs. P shown in Figure 7.6.9.

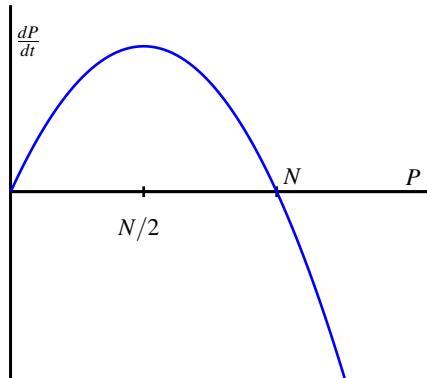


Figure 7.6.9: Plot of $\frac{dP}{dt}$ vs. P .

- At what value of P is the rate of change greatest?
- Consider the model for the earth's population that we created. At what value of P is the rate of change greatest? How does that compare to the population in recent years?
- According to the model we developed, what will the population be in the year 2100?
- According to the model we developed, when will the population reach 9 billion?
- Now consider the general solution to the general logistic initial value problem that we found, given by

$$P(t) = \frac{N}{\left(\frac{N-P_0}{P_0}\right) e^{-kNt} + 1}.$$

Verify algebraically that $P(0) = P_0$ and that $\lim_{t \rightarrow \infty} P(t) = N$.

Hint.

- Small hints for each of the prompts above.

Answer.

- When $P = \frac{N}{2}$.
- When the population is 6.125 billion.

- c. $P = \frac{12.5}{1.0546e^{-0.025t} + 1}$; $P(100) = 11.504$ billion.
- d. $t = \frac{1}{-0.025} \ln \left(\frac{\left(\frac{12.5}{9} - 1 \right)}{1.0546} \right) \approx 39.9049$ (so in about year 2040).
- e. $\lim_{t \rightarrow \infty} P(t) = N$.

Solution. Consider the logistic equation $\frac{dP}{dt} = kP(N - P)$.

- a. The maximum value for $\frac{dP}{dt}$ will occur when $P = \frac{N}{2}$ since this is the highest point on the graph.
- b. The model we created for the population of the earth was $\frac{dP}{dt} = P(0.025 - 0.002P)$, which can be rewritten as $\frac{dP}{dt} = 0.002P(12.5 - P)$. So for this model, the maximum rate of change for the earth's population will occur when the population is 6.125 billion. This value is slightly less than the earth's population in recent years.
- c. So for the logistic model, we have $k = 0.002$, $N = 12.5$. Using the year 2000 as $t = 0$, the initial population is 6.084. So the solution for the differential equation in part (b) is

$$P = \frac{12.5}{1.0546e^{-0.025t} + 1}.$$

For the year 2100, we use $t = 100$ and this model predicts that in the year 2100, the earth's population will be 11.504 billion.

- d. To determine when this model predicts that the earth's population will be 9 billion, we solve the equation

$$\frac{12.5}{1.0546e^{-0.025t} + 1} = 9$$

for t . Multiplying both sides by $(1.0546e^{-0.025t} + 1)$, we have

$$12.5 = 9(1.0546e^{-0.025t} + 1).$$

Dividing by 9 then subtracting 1 from both sides, $1.0546e^{-0.025t} = \left(\frac{12.5}{9} - 1 \right)$. Thus, $e^{-0.025t} = \frac{\left(\frac{12.5}{9} - 1 \right)}{1.0546}$. Using logarithms in the usual way, $t = \frac{1}{-0.025} \ln \left(\frac{\left(\frac{12.5}{9} - 1 \right)}{1.0546} \right) \approx 39.9049$.

The earth's population will reach 9 billion in about 2040.

- e. The solution for the general logistic initial value problem is

$$P(t) = \frac{N}{\left(\frac{N-P_0}{P_0} \right) e^{-kNt} + 1}.$$

For this solution, we see that

$$\begin{aligned} P(0) &= \frac{N}{\left(\frac{N-P_0}{P_0} \right) + 1} \\ &= \frac{N}{\frac{N-P_0+P_0}{P_0}} \end{aligned}$$

$$\begin{aligned}
 &= \frac{NP_0}{N} \\
 &= P_0
 \end{aligned}$$

To determine $\lim_{t \rightarrow \infty} P(t)$, we first note that since $-kN < 0$, $\lim_{t \rightarrow \infty} e^{-kNt} = 0$. Hence,

$$\lim_{t \rightarrow \infty} \left[\left(\frac{N - P_0}{P_0} \right) e^{-kNt} + 1 \right] = 1,$$

and therefore,

$$\lim_{t \rightarrow \infty} P(t) = \lim_{t \rightarrow \infty} \frac{N}{\left(\frac{N - P_0}{P_0} \right) e^{-kNt} + 1} = \frac{N}{1} = N.$$

7.6.3 Summary

- If we assume that the rate of growth of a population is proportional to the population, we are led to a model in which the population grows without bound and at a rate that grows without bound.
- By assuming that the per capita growth rate decreases as the population grows, we are led to the logistic model of population growth, which predicts that the population will eventually stabilize at the carrying capacity.

7.6.4 Exercises

1. (Analyzing a logistic equation) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Analyzing a logistic model) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Finding a logistic function for an infection model) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Analyzing a population growth model) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. The logistic equation may be used to model how a rumor spreads through a group of people. Suppose that $p(t)$ is the fraction of people that have heard the rumor on day t . The equation

$$\frac{dp}{dt} = 0.2p(1 - p)$$

describes how p changes. Suppose initially that one-tenth of the people have heard the rumor; that is, $p(0) = 0.1$.

- a. What happens to $p(t)$ after a very long time?
- b. Determine a formula for the function $p(t)$.
- c. At what time is p changing most rapidly?
- d. How long does it take before 80% of the people have heard the rumor?

Answer.

- a. $p(t) \rightarrow 1$ as $t \rightarrow \infty$ provided $p(0) > 0$.
- b. $p(t) = \frac{1}{9e^{-0.2t} + 1}$.
- c. $t = -5 \ln(1/9) \approx 10.986$ days.
- d. $t = -5 \ln(0.2/9) \approx 19.033$ days.

Solution.

- a. The equilibrium solutions for this logistic equation are $p = 0$ and $p = 1$. Since $p = 1$ is the carrying capacity and thus a stable equilibrium, $p(t) \rightarrow 1$ as $t \rightarrow \infty$ provided $p(0) > 0$.
- b. We know that the general solution to the logistic equation $\frac{dp}{dt} = kP(N - P)$, $P(0) = P_0$ is

$$P(t) = \frac{N}{\left(\frac{N-P_0}{P_0}\right) e^{-kNt} + 1}.$$

In the context of this problem, we know that $k = 0.2$, $N = 1$, and $p(0) = 0.1$ and therefore the solution is

$$p(t) = \frac{1}{\left(\frac{1-0.1}{0.1}\right) e^{-0.2t} + 1} = \frac{1}{9e^{-0.2t} + 1}.$$

- c. We know that p is changing most rapidly when its derivative, p' , is largest. Recalling the original differential equation, we know that

$$p' = 0.2p(1 - p)$$

which shows that p' is a quadratic function of p . That quadratic function has zeros at $p = 0$ and $p = 1$, so the vertex of this concave down quadratic function lies at $p = \frac{1}{2}$, and at this value of p the rate of change of p is largest. Thus, we solve the equation $p(t) = 0.5$ to find the time at which the population is increasing fastest. We see that

$$0.5 = \frac{1}{9e^{-0.2t} + 1}$$

so

$$9e^{-0.2t} + 1 = 2$$

and $e^{-0.2t} = 1/9$. It follows that $t = -5 \ln(1/9) \approx 10.986$ days.

- d. Since $p(t)$ tells us the percentage of people who've heard the rumor, we need to solve the equation $p(t) = 0.8$. Thus we have

$$0.8 = \frac{1}{9e^{-0.2t} + 1}$$

so

$$9e^{-0.2t} + 1 = 1.2$$

and $e^{-0.2t} = 0.2/9$. It follows that $t = -5 \ln(0.2/9) \approx 19.033$ days.

- 6. Suppose that $b(t)$ measures the number of bacteria living in a colony in a Petri dish, where b is measured in thousands and t is measured in days. One day, you measure that there are 6,000 bacteria and the per capita growth rate is 3. A few days later, you measure that there are 9,000 bacteria and the per capita growth rate is 2.

- Assume that the per capita growth rate $\frac{db}{dt}$ is a linear function of b . Use the measurements to find this function and write a logistic equation to describe $\frac{db}{dt}$.
- What is the carrying capacity for the bacteria?
- At what population is the number of bacteria increasing most rapidly?
- If there are initially 1,000 bacteria, how long will it take to reach 80% of the carrying capacity?

Answer.

- $\frac{db}{dt} = \frac{1}{3000}b(15000 - b)$
- $b = 15000$.
- When $b = 7500$.
- $t = -\frac{1}{5} \ln(1/70) \approx 0.8497$ days.

Solution.

- Let y be the per capita growth rate as a function of b . Using the given data in the problem statement, we know that this linear function passes through the points $(6000, 3)$ and $(9000, 2)$. The equation of this line is thus

$$y - 3 = -\frac{1}{3000}(b - 6000)$$

so $y = (3 + 2) - \frac{1}{3000}b = 5 - \frac{1}{3000}b$. Since the logistic equation comes from taking the product of the population at time t with its per capita growth rate at that time, we find that

$$\frac{db}{dt} = b \left(5 - \frac{1}{3000}b \right) = \frac{1}{3000}b(15000 - b)$$

- From the form of the logistic equation, we see that the carrying capacity for the bacteria is $b = 15000$, which is a stable equilibrium for the population.
- The bacteria is increasing most rapidly when $\frac{db}{dt}$ is greatest. Since

$$\frac{db}{dt} = \frac{1}{3000}b(15000 - b)$$

is a quadratic function of b , $\frac{db}{dt}$ is maximized at the vertex of this concave down parabola, which occurs halfway between its zeros: $b = 0$ and $b = 15000$. Thus, when $b = 7500$, the population is growing fastest. To find the exact t -value when this occurs, we'd need to know an initial condition like in the next part of the problem.

- Using $b(0) = 1000$, we solve the logistic equation IVP with

$$\frac{db}{dt} = \frac{1}{3000}b(15000 - b).$$

Using $N = 15000$, $k = \frac{1}{3000}$, and $b_0 = 1000$, we know from the standard form of the solution to the logistic equation that

$$b(t) = \frac{15000}{\left(\frac{15000-1000}{1000}\right)e^{-\frac{1}{3000}t} + 1} = \frac{15000}{14e^{-\frac{1}{3000}t} + 1}.$$

To find how long it takes the bacteria to reach 80% of the carrying capacity, we solve the equation

$$0.8 \cdot 15 = \frac{15000}{14e^{-5t} + 1}.$$

That equation implies that

$$0.8 = \frac{1}{14e^{-5t} + 1},$$

so $14e^{-5t} + 1 = 1.2$ and thus $e^{-5t} = 0.2/14 = 1/70$. It follows that $t = -\frac{1}{5} \ln(1/70) \approx 0.8497$ days.

7. Suppose that the population of a species of fish is controlled by the logistic equation

$$\frac{dP}{dt} = 0.1P(10 - P),$$

where P is measured in thousands of fish and t is measured in years.

- What is the carrying capacity of this population?
- Suppose that a long time has passed and that the fish population is stable at the carrying capacity. At this time, humans begin harvesting 20% of the fish every year. Modify the differential equation by adding a term to incorporate the harvesting of fish.
- What is the new carrying capacity?
- What will the fish population be one year after the harvesting begins?
- How long will it take for the population to be within 10% of the carrying capacity?

Answer.

- 10000 fish.
- $\frac{dP}{dt} = 0.1P(10 - P) - 0.2P$.
- 8000 fish.
- $P(1) \approx 8.7899$ thousand fish.
- $t = -1.2 \ln(5/11) \approx 0.9461$ years.

Solution.

- From the form of the given logistic equation, we can see the carrying capacity is 10000 fish.
- In this situation, the initial condition is $P(0) = 10$, and by removing 20% of the population per year, we subtract $0.2P$ from the rate that defines the differential equation, thus finding

$$\frac{dP}{dt} = 0.1P(10 - P) - 0.2P.$$

- Expanding and rewriting the differential equation in (b), we see that

$$\frac{dP}{dt} = P - 0.1P^2 - 0.2P = 0.8P - 0.1P^2 = 0.1P(8 - P).$$

This, too, is a logistic equation, with carrying capacity 8000 fish.

- d. We can solve this logistic equation in the usual way (keeping in mind the initial condition $P(0) = 10$, and noting that the carrying capacity is 8):

$$P(t) = \frac{8}{\frac{8-10}{10}e^{-0.18t} + 1} = \frac{8}{-0.2e^{-0.8t} + 1}$$

After one year, the population of fish will thus be

$$P(1) \approx 8.7899$$

thousand fish.

- e. To be within 10% of the carrying capacity, we need $0.9(8) \leq P(t) \leq 1.1(8)$. Since the initial condition is larger than the carrying capacity, we want to know when $P(t) = 1.1(8) = 8.8$. Thus we solve the equation

$$8.8 = \frac{8}{-0.2e^{-0.8t} + 1},$$

which implies $-0.2e^{-0.8t} + 1 = 8/8.8 = 10/11$. Thus, $e^{-0.8t} = 5/11$, so $t = -1.2 \ln(5/11) \approx 0.9461$ years.

Chapter 8

Sequences and Series

8.1 Sequences

Motivating Questions

- What is a sequence?
 - What does it mean for a sequence to converge?
 - What does it mean for a sequence to diverge?

We encounter sequences every day. Your monthly utility payments, the annual interest you earn on investments, the amount you spend on groceries each week; all are examples of sequences. Other sequences with which you may be familiar include the Fibonacci sequence $1, 1, 2, 3, 5, 8, \dots$, where each term is the sum of the two preceding terms, and the triangular numbers $1, 3, 6, 10, 15, 21, 28, 36, 45, 55, \dots$, the number of vertices in the triangles shown in Figure 8.1.1.

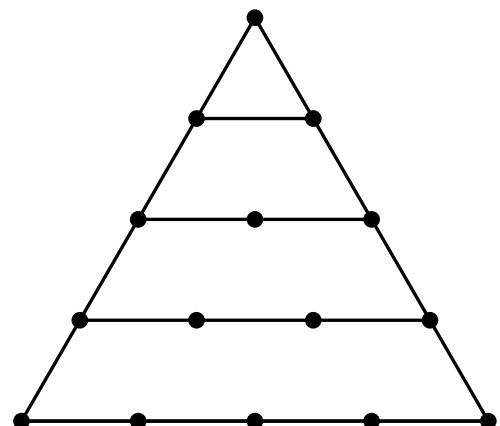


Figure 8.1.1: Triangular numbers

Sequences of integers are of such interest to mathematicians and others that they have a

journal¹ devoted to them and an on-line encyclopedia² that catalogs a huge number of integer sequences and their connections. Sequences are also used in digital recordings and images.

Our studies in calculus have dealt with continuous continuous functions. Sequences model *discrete* instead of continuous information. We will study ways to represent and work with discrete information in this chapter as we investigate *sequences* and *series*, and ultimately see key connections between the discrete and continuous.

Example 8.1.2. Suppose you receive \$5000 through an inheritance. You decide to invest this money into a fund that pays 8% annually, compounded monthly. That means that each month your investment earns $\frac{0.08}{12} \cdot P$ additional dollars, where P is your principal balance at the start of the month. So in the first month your investment earns

$$5000 \left(\frac{0.08}{12} \right)$$

or \$33.33. If you reinvest this money, you will then have \$5033.33 in your account at the end of the first month. From this point on, assume that you reinvest all of the interest you earn.

- How much interest will you earn in the second month? How much money will you have in your account at the end of the second month?
- Complete Table 8.1.3 to determine the interest earned and total amount of money in this investment each month for one year.

Month	Interest earned	Total amount of money in the account
0	\$0.00	\$5000.00
1	\$33.33	\$5033.33
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		

Table 8.1.3: Interest

¹The *Journal of Integer Sequences* at <http://www.cs.uwaterloo.ca/journals/JIS/>

²The On-Line Encyclopedia of Integer Sequences at <http://oeis.org/>

- c. As we will see later, the amount of money P_n in the account after month n is given by

$$P_n = 5000 \left(1 + \frac{0.08}{12}\right)^n.$$

Use this formula to check your calculations in Table 8.1.3. Then find the amount of money in the account after 5 years.

- d. How many years will it be before the account has doubled in value to \$10000?

8.1.1 Sequences

As Example 8.1.2 illustrates, many discrete phenomena can be represented as lists of numbers (like the amount of money in an account over a period of months). We call any such list a *sequence*. A sequence is nothing more than list of terms in some order. We often list the entries of the sequence with subscripts,

$$s_1, s_2, \dots, s_n \dots,$$

where the subscript denotes the position of the entry in the sequence.

Definition 8.1.4. A sequence is a list of terms s_1, s_2, s_3, \dots in a specified order.

We can think of a sequence as a function f whose domain is the set of positive integers where $f(n) = s_n$ for each positive integer n . This alternative view will be useful in many situations.

We often denote the sequence

$$s_1, s_2, s_3, \dots$$

by $\{s_n\}$. The value s_n (alternatively $s(n)$) is called the n th *term* in the sequence. If the terms are all 0 after some fixed value of n , we say the sequence is finite. Otherwise the sequence is infinite. With infinite sequences, we are often interested in their end behavior and the idea of *convergent* sequences.

Example 8.1.5.

- a. Let s_n be the n th term in the sequence $1, 2, 3, \dots$. Find a formula for s_n and use appropriate technological tools to draw a graph of entries in this sequence by plotting points of the form (n, s_n) for some values of n . Most graphing calculators can plot sequences; directions follow for the TI-84.

- In the MODEmenu, highlight SEQ in the FUNClue and press ENTER.
- In the Y=menu, you will now see lines to enter sequences. Enter a value for nMin (where the sequence starts), a function for u(n) (the n th term in the sequence), and the value of u_{n Min}.
- Set your window coordinates (this involves choosing limits for n as well as the window coordinates XMin, XMax, YMin, and YMax).
- The GRAPHkey will draw a plot of your sequence.

Using your knowledge of limits of continuous functions as $x \rightarrow \infty$, decide if this sequence $\{s_n\}$ has a limit as $n \rightarrow \infty$. Explain your reasoning.

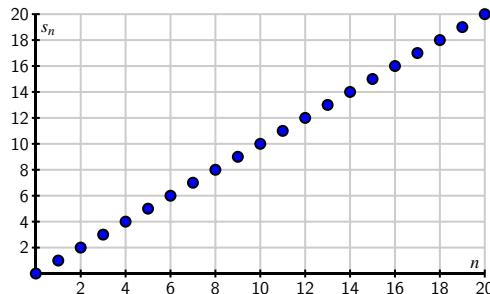
- Let s_n be the n th term in the sequence $1, \frac{1}{2}, \frac{1}{3}, \dots$. Find a formula for s_n . Draw a graph of some points in this sequence. Using your knowledge of limits of continuous functions as $x \rightarrow \infty$, decide if this sequence $\{s_n\}$ has a limit as $n \rightarrow \infty$. Explain your reasoning.
- Let s_n be the n th term in the sequence $2, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \dots$. Find a formula for s_n . Using your knowledge of limits of continuous functions as $x \rightarrow \infty$, decide if this sequence $\{s_n\}$ has a limit as $n \rightarrow \infty$. Explain your reasoning.

Hint.

- When $n = 1$, $s_n = 1$; when $n = 2$, $s_n = 2$.
- Think about the value of $\frac{1}{n}$.
- Note that the numerator of each term of the sequence is one more than the denominator of the term.

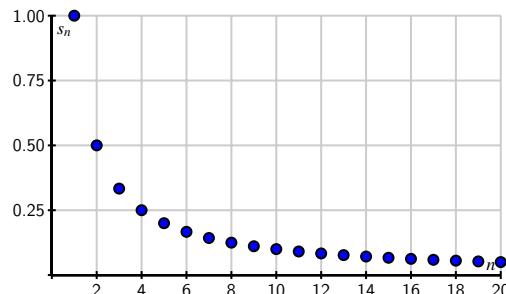
Answer.

- $s_n = n$. A plot of the first 50 points in the sequence is shown here.



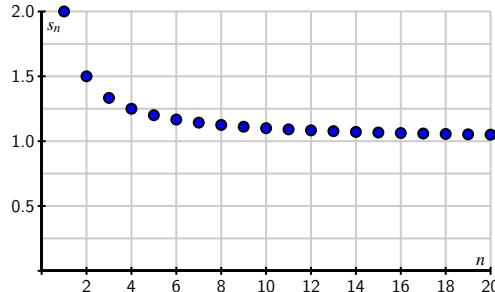
This sequence does not have a limit as n goes to infinity.

- s_n is $s_n = \frac{1}{n}$. A plot of the first 50 points in the sequence is shown here.



This sequence has a limit of 0 as n goes to infinity.

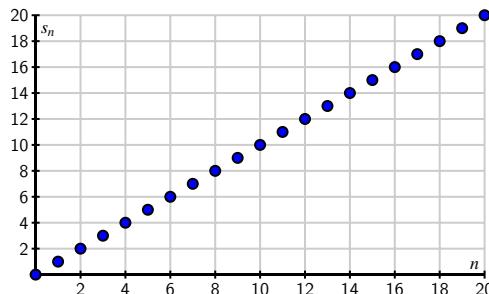
- c. $s_n = \frac{n+1}{n}$. A plot of the first 50 points in the sequence is shown here.



This sequence has a limit of 1 as n goes to infinity.

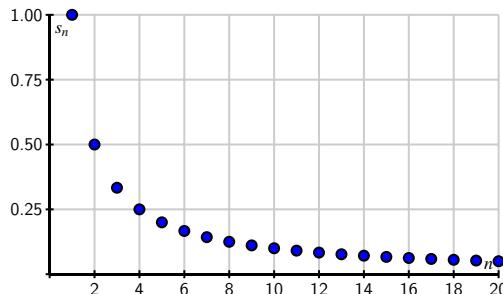
Solution.

- a. By observation we see that a formula for s_n is $s_n = n$. A plot of the first 50 points in the sequence is shown here.



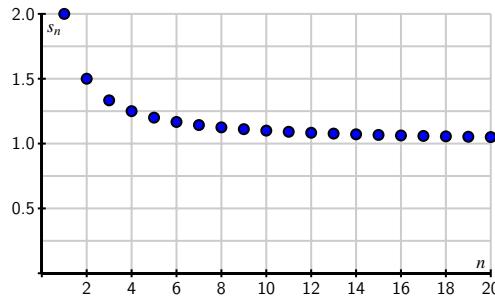
We recalling that a function f has a limit L at infinity if we can make the values of $f(x)$ as large as we want by choosing x as large as we need. Since we can make the values of n in our sequence as large as we want by choosing n to be as large as we need, we suspect that this sequence does not have a limit as n goes to infinity.

- b. By observation we see that a formula for s_n is $s_n = \frac{1}{n}$. A plot of the first 50 points in the sequence is shown here.



Since we can make the values of $\frac{1}{n}$ in our sequence as close to 0 as we want by choosing n to be as large as we need, we suspect that this sequence has a limit of 0 as n goes to infinity.

- c. Since the numerator is always 1 more than the denominator, a formula for s_n is $s_n = \frac{n+1}{n}$. A plot of the first 50 points in the sequence is shown here.



Since we can make the values of $\frac{n+1}{n}$ in our sequence as close to 1 as we want by choosing n to be as large as we need, we suspect that this sequence has a limit of 1 as n goes to infinity.

Next we formalize the ideas from Example 8.1.5.

Example 8.1.6.

- Recall our earlier work with limits involving infinity in Section 3.6. State clearly what it means for a continuous function f to have a limit L as $x \rightarrow \infty$.
- Given that an infinite sequence of real numbers is a function from the integers to the real numbers, apply the idea from part (a) to explain what you think it means for a sequence $\{s_n\}$ to have a limit as $n \rightarrow \infty$.
- Based on your response to the part (b), decide if the sequence $\{\frac{1+n}{2+n}\}$ has a limit as $n \rightarrow \infty$. If so, what is the limit? If not, why not?

Hint.

- A function f has limit L as $x \rightarrow \infty$ if we can make $f(x)$ as close to L as we like by
- Think about making s_n as close as you want to L for large values of n .
- Compare $f(x) = \frac{1+x}{2+x}$ as $x \rightarrow \infty$.

Solution.

- A continuous function f has a limit L as the independent variable x goes to infinity if we can make the values of $f(x)$ as close to L as we want by choosing x as large as we need.
- We expect that a sequence $\{s_n\}$ will have a limit L as n goes to infinity if we can make the entries s_n in the sequence as close to L as we want by choosing n as large as we need.

- c. As n gets large, the constant terms become infinitesimally small compared to n and so $\frac{1+n}{2+n}$ looks like $\frac{n}{n}$ or 1 for large n . So the sequence $\{\frac{1+n}{2+n}\}$ has a limit of 1 at infinity.

In Activities 8.1.5 and 8.1.6 we investigated a sequence $\{s_n\}$ that has a limit as n goes to infinity. More formally, we make the following definition.

Definition 8.1.7. A sequence $\{s_n\}$ **converges** or is a **convergent sequence** provided that there is a number L such that we can make s_n as close to L as we want by taking n sufficiently large. In this situation, we call L the *limit* of the convergent sequence and write

$$\lim_{n \rightarrow \infty} s_n = L.$$

If the sequence $\{s_n\}$ does not converge, we say that the sequence $\{s_n\}$ **diverges**.

The idea of sequence having a limit as $n \rightarrow \infty$ is the same as the idea of a continuous function having a limit as $x \rightarrow \infty$. The only difference is that sequences are discrete instead of continuous.

Example 8.1.8. Use graphical and/or algebraic methods to determine whether each of the following sequences converges or diverges.

- $\{\frac{1+2n}{3n-2}\}$
- $\{\frac{5+3^n}{10+2^n}\}$
- $\{\frac{10^n}{n!}\}$ (where $!$ is the *factorial* symbol and $n! = n(n-1)(n-2)\cdots(2)(1)$ for any positive integer n (as convention we define $0!$ to be 1)).

Hint.

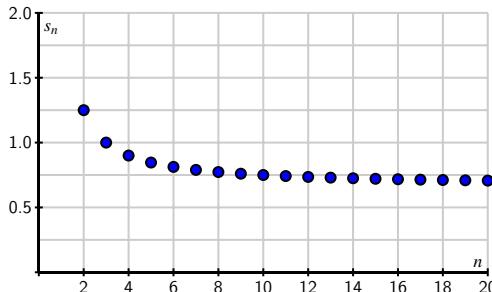
- Multiply the numerator and denominator each by $\frac{1}{n}$.
- Compare $\frac{3^n}{2^n}$.
- Plot the sequence and think about what the graph suggests.

Answer.

- The sequence $\{\frac{1+2n}{3n-2}\}$ converges to $\frac{2}{3}$.
- The sequence $\{\frac{5+3^n}{10+2^n}\}$ diverges to infinity.
- $\frac{10^n}{n!} \rightarrow 0$ as $n \rightarrow \infty$.

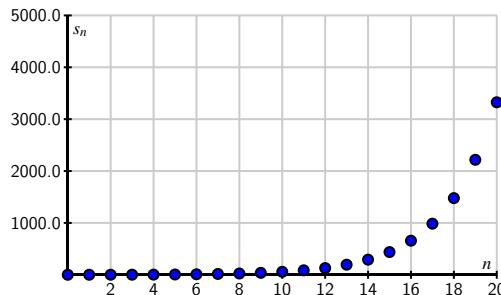
Solution.

- A plot of the first 50 terms of the sequence $\{\frac{1+2n}{3n-2}\}$ is shown here.



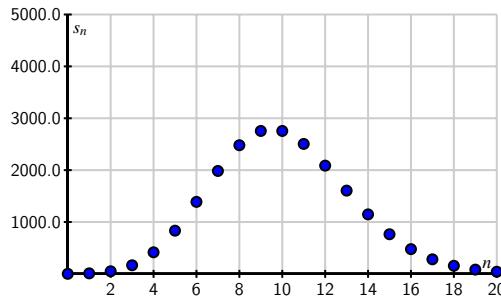
The plot implies that the sequence has a limit between 0.5 and 1. For large n the $2n$ term dominates the numerator and the $3n$ term the denominator. So $\frac{1+2n}{3n-2}$ looks like $\frac{2n}{3n} = \frac{2}{3}$ when n is big. So the sequence $\{\frac{1+2n}{3n-2}\}$ converges to $\frac{2}{3}$.

- b. A plot of the first 50 terms of the sequence $\{\frac{5+3^n}{10+2^n}\}$ is shown here.



The plot implies that the sequence does not have a limit as n goes to infinity. For large n the 3^n term dominates the numerator and the 2^n term the denominator. So $\frac{5+3^n}{10+2^n}$ looks like $\frac{3^n}{2^n} = \left(\frac{3}{2}\right)^n$ when n is big. Since $\frac{3}{2} > 1$, the sequence $\{\frac{5+3^n}{10+2^n}\}$ diverges to infinity.

- c. A plot of the first 50 terms of the sequence $\{\frac{10^n}{n!}\}$ is shown here.



Initially, it looks as though the terms increase without bound, but beginning at about $n = 10$ the factorial in the denominator dominates the numerator. Notice that

$$\frac{10^n}{n!} = \frac{10 \times 10 \times 10 \times \cdots \times 10}{1 \times 2 \times 3 \times \cdots \times n}$$

When $n > 20$, we have that $\frac{10}{n} < \frac{1}{2}$ and

$$\frac{10^n}{n!} = \left(\frac{10 \times 10 \times 10 \times \cdots \times 10}{1 \times 2 \times 3 \times \cdots \times 20} \right) \left(\frac{10 \times 10 \times 10 \times \cdots \times 10}{21 \times 22 \times 23 \times \cdots \times n} \right)$$

$$\begin{aligned}
 &= \left(\frac{10^{20}}{20!} \right) \left(\frac{10}{21} \right) \left(\frac{10}{22} \right) \cdots \left(\frac{10}{n} \right) \\
 &< \left(\frac{10^{20}}{20!} \right) \left(\frac{1}{2} \right)^{n-20}.
 \end{aligned}$$

Since $\frac{1}{2} < 1$, the term $\left(\frac{1}{2} \right)^{n-20}$ goes to 0 as n goes to infinity. The fact that $\frac{10^{20}}{20!}$ is a constant means that $\frac{10^n}{n!} \rightarrow 0$ as $n \rightarrow \infty$.

8.1.2 Summary

- A sequence is a list of objects in a specified order. We will typically work with sequences of real numbers. We can think of a sequence as a function from the positive integers to the set of real numbers.
- A sequence $\{s_n\}$ of real numbers converges to a number L if we can make every value of s_k for $k \geq n$ as close as we want to L by choosing n sufficiently large.
- A sequence diverges if it does not converge.

8.1.3 Exercises

1. (Limits of five sequences) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Formula for a sequence, given first terms) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Divergent or convergent sequences) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Terms of a sequence from sampling a signal) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. Finding limits of convergent sequences can be a challenge. However, there is a useful tool we can adapt from our study of limits of continuous functions at infinity to use to find limits of sequences. We illustrate in this exercise with the example of the sequence

$$\frac{\ln(n)}{n}.$$

- a. Calculate the first 10 terms of this sequence. Based on these calculations, do you think the sequence converges or diverges? Why?
- b. For this sequence, there is a corresponding continuous function f defined by

$$f(x) = \frac{\ln(x)}{x}.$$

Draw the graph of $f(x)$ on the interval $[0, 10]$ and then plot the entries of the sequence on the graph. What conclusion do you think we can draw about the sequence $\left\{ \frac{\ln(n)}{n} \right\}$ if $\lim_{x \rightarrow \infty} f(x) = L$? Explain.

- c. Note that $f(x)$ has the indeterminate form $\frac{\infty}{\infty}$ as x goes to infinity. What idea from differential calculus can we use to calculate $\lim_{x \rightarrow \infty} f(x)$? Use this method to find $\lim_{x \rightarrow \infty} f(x)$. What, then, is $\lim_{n \rightarrow \infty} \frac{\ln(n)}{n}$?

Answer.

- a. Unclear whether it converges or diverges.

n	1	2	3	4	5
$\frac{\ln(n)}{n}$	0	0.3466	0.3663	0.3466	0.3218
t	6	7	8	9	10
$\frac{\ln(n)}{n}$	0.2987	0.2781	0.2599	0.2442	0.2303

- b. If $\lim_{x \rightarrow \infty} f(x) = L$, then $\lim_{n \rightarrow \infty} \frac{\ln(n)}{n} = L$ as well.

- c.

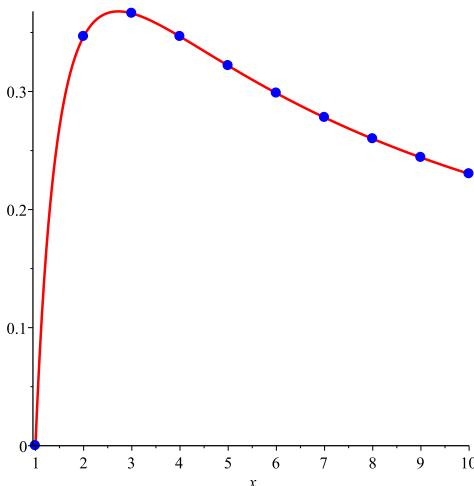
$$\lim_{n \rightarrow \infty} \frac{\ln(n)}{n} = \lim_{x \rightarrow \infty} \frac{\ln(x)}{x} = 0.$$

Solution.

- a. Decimal approximations of the first 10 terms of this sequence are given in the following table. The numbers in this sequence originally increase, but then seem to be decreasing. It looks as though the sequence might converge, but if so it isn't clear to what value the sequence converges.

n	1	2	3	4	5
$\frac{\ln(n)}{n}$	0	0.3466	0.3663	0.3466	0.3218
t	6	7	8	9	10
$\frac{\ln(n)}{n}$	0.2987	0.2781	0.2599	0.2442	0.2303

- b. The graph of f on the interval $[1, 10]$ and the graph of the sequence $\left\{ \frac{\ln(n)}{n} \right\}$ for n from 1 to 10 are shown in the figure below. Since the graph of the sequence follows the graph of the function f , and since f is a strictly decreasing function, we should expect that if $\lim_{x \rightarrow \infty} f(x) = L$, then $\lim_{n \rightarrow \infty} \frac{\ln(n)}{n} = L$ as well.



- c. Since we have the indeterminate form $\frac{\infty}{\infty}$, we can apply L'Hopital's Rule to determine the limit. In this case, we have

$$\lim_{x \rightarrow \infty} \frac{\ln(x)}{x} = \lim_{x \rightarrow \infty} \frac{1}{x} = 0.$$

So we should expect that

$$\lim_{n \rightarrow \infty} \frac{\ln(n)}{n} = \lim_{x \rightarrow \infty} \frac{\ln(x)}{x} = 0.$$

6. We return to the example begun in Preview Activity 8.1.2 to see how to derive the formula for the amount of money in an account at a given time. We do this in a general setting. Suppose you invest P dollars (called the *principal*) in an account paying $r\%$ interest compounded monthly. In the first month you will receive $\frac{r}{12}$ (here r is in decimal form; e.g., if we have 8% interest, we write $\frac{0.08}{12}$) of the principal P in interest, so you earn

$$P \left(\frac{r}{12} \right)$$

dollars in interest. Assume that you reinvest all interest. Then at the end of the first month your account will contain the original principal P plus the interest, or a total of

$$P_1 = P + P \left(\frac{r}{12} \right) = P \left(1 + \frac{r}{12} \right)$$

dollars.

- a. Given that your principal is now P_1 dollars, how much interest will you earn in the second month? If P_2 is the total amount of money in your account at the end of the second month, explain why

$$P_2 = P_1 \left(1 + \frac{r}{12} \right) = P \left(1 + \frac{r}{12} \right)^2.$$

- b. Find a formula for P_3 , the total amount of money in the account at the end of the third month in terms of the original investment P .
 c. There is a pattern to these calculations. Let P_n the total amount of money in the account at the end of the third month in terms of the original investment P . Find a formula for P_n .

Answer.

- $P_1 \left(\frac{r}{12} \right)$ in interest in the second month; at the end of the second month, $P_2 = P \left(1 + \frac{r}{12} \right)^2$.
- $P_3 = P \left(1 + \frac{r}{12} \right)^3$.
- $P_n = P \left(1 + \frac{r}{12} \right)^n$ is a pattern to these calculations.

Solution.

- a. The principal of P_1 dollars will earn $P_1 \left(\frac{r}{12} \right)$ in interest in the second month. Thus, at the end of the second month, the total amount P_2 in your account will be

$$P_2 = P_1 + P_1 \left(\frac{r}{12} \right).$$

Factoring and recalling that $P_1 = P \left(1 + \frac{r}{12}\right)$

$$P_2 = P_1 \left(1 + \frac{r}{12}\right) = P \left(1 + \frac{r}{12}\right)^2.$$

- b. The principal of P_2 dollars will earn $P_2 \left(\frac{r}{12}\right)$ in interest in the third month. Thus, at the end of the third month, the total amount P_3 in your account will be

$$P_3 = P_2 + P_2 \left(\frac{r}{12}\right).$$

Factoring and recalling that $P_2 = P \left(1 + \frac{r}{12}\right)^2$

$$P_3 = P_2 \left(1 + \frac{r}{12}\right) = P \left(1 + \frac{r}{12}\right)^3.$$

- c. Based on the pattern we observe, continuing the same line of reasoning we find that

$$P_n = P_{n-1} \left(1 + \frac{r}{12}\right) = P \left(1 + \frac{r}{12}\right)^n$$

is a pattern to these calculations.

7. Sequences have many applications in mathematics and the sciences. In a recent paper³ the authors write

The incretin hormone glucagon-like peptide-1 (GLP-1) is capable of ameliorating glucose-dependent insulin secretion in subjects with diabetes. However, its very short half-life (1.5-5 min) in plasma represents a major limitation for its use in the clinical setting.

The half-life of GLP-1 is the time it takes for half of the hormone to decay in its medium. For this exercise, assume the half-life of GLP-1 is 5 minutes. So if A is the amount of GLP-1 in plasma at some time t , then only $\frac{A}{2}$ of the hormone will be present after $t + 5$ minutes. Suppose $A_0 = 100$ grams of the hormone are initially present in plasma.

- Let A_1 be the amount of GLP-1 present after 5 minutes. Find the value of A_1 .
- Let A_2 be the amount of GLP-1 present after 10 minutes. Find the value of A_2 .
- Let A_3 be the amount of GLP-1 present after 15 minutes. Find the value of A_3 .
- Let A_4 be the amount of GLP-1 present after 20 minutes. Find the value of A_4 .
- Let A_n be the amount of GLP-1 present after $5n$ minutes. Find a formula for A_n .
- Does the sequence $\{A_n\}$ converge or diverge? If the sequence converges, find its limit and explain why this value makes sense in the context of this problem.
- Determine the number of minutes it takes until the amount of GLP-1 in plasma is 1 gram.

Answer. Exercise Answer

Solution.

³Hui H, Farilla L, Merkel P, Perfetti R. The short half-life of glucagon-like peptide-1 in plasma does not reflect its long-lasting beneficial effects, *Eur J Endocrinol* 2002 Jun;146(6):863-9.

- a. Since one half-life has passed at $t = 5$, $A_1 = \frac{1}{2}A_0 = \frac{1}{2}(100)$.
- b. Since another half-life has passed at $t = 10$, $A_2 = \frac{1}{2}A_1 = \frac{1}{2}\left(\frac{1}{2}(100)\right) = \left(\frac{1}{2}\right)^2 \cdot 100$.
- c. Yet another half-life has passed at $t = 15$, so $A_3 = \frac{1}{2}A_2 = \frac{1}{2}\left(\left(\frac{1}{2}\right)^2 \cdot 100\right) = \left(\frac{1}{2}\right)^3 \cdot 100$.
- d. Continuing, $A_4 = \frac{1}{2}A_3 = \frac{1}{2}\left(\left(\frac{1}{2}\right)^3 \cdot 100\right) = \left(\frac{1}{2}\right)^4 \cdot 100$.
- e. After $5n$ minutes, n half-lives have elapsed, and thus we find that

$$A_n = \left(\frac{1}{2}\right)^n \cdot 100.$$

- f. From the behavior of exponential functions, we know that if $0 < a < 1$, then $a^x \rightarrow 0$ as $x \rightarrow \infty$. It follows that $\left(\frac{1}{2}\right)^n \rightarrow 0$ as $n \rightarrow \infty$, and therefore $A_n \rightarrow 0$ as time goes on. This makes sense because the human body will consume and change the drug in the body and eliminate it unless additional doses are taken.
- g. To determine the time at which 1 gram of the hormone remains in the body, we solve the equation

$$1 = \left(\frac{1}{2}\right)^n \cdot 100.$$

This implies that $\left(\frac{1}{2}\right)^n = 0.01$, and thus $n = \frac{\ln(0.01)}{\ln(0.5)} \approx 6.6439$. Thus, it takes about 6.6439 half-lives to elapse to get down to 1 gram remaining, or $5 \cdot 6.6439 = 33.2193$ minutes.

8. Continuous data is the basis for analog information, like music stored on old cassette tapes or vinyl records. A digital signal like on a CD or MP3 file is obtained by sampling an analog signal at some regular time interval and storing that information. For example, the sampling rate of a compact disk is 44,100 samples per second. So a digital recording is only an approximation of the actual analog information. Digital information can be manipulated in many useful ways that allow for, among other things, noisy signals to be cleaned up and large collections of information to be compressed and stored in much smaller space. While we won't investigate these techniques in this chapter, this exercise is intended to give an idea of the importance of discrete (digital) techniques.

Let f be the continuous function defined by $f(x) = \sin(4x)$ on the interval $[0, 10]$. A graph of f is shown in Figure 8.1.9.

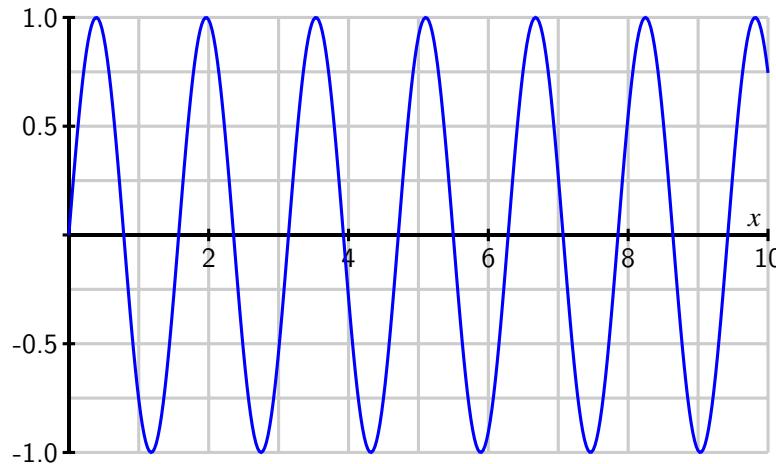


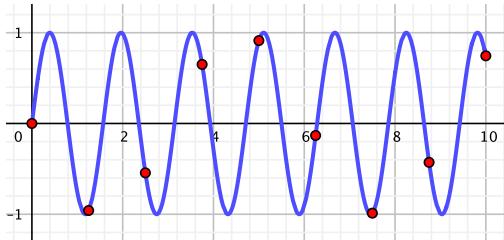
Figure 8.1.9: The graph of $f(x) = \sin(4x)$ on the interval $[0, 10]$

We approximate f by *sampling*, that is by partitioning the interval $[0, 10]$ into uniform subintervals and recording the values of f at the endpoints.

- Ineffective sampling can lead to several problems in reproducing the original signal. As an example, partition the interval $[0, 10]$ into 8 equal length subintervals and create a list of points (the *sample*) using the endpoints of each subinterval. Plot your sample on graph of f in Figure Figure 8.1.9. What can you say about the period of your sample as compared to the period of the original function?
- The sampling rate is the number of samples of a signal taken per second. As the part (a) illustrates, sampling at too small a rate can cause serious problems with reproducing the original signal (this problem of inefficient sampling leading to an inaccurate approximation is called *aliasing*). There is an elegant theorem called the Nyquist-Shannon Sampling Theorem that says that human perception is limited, which allows that replacement of a continuous signal with a digital one without any perceived loss of information. This theorem also provides the *lowest* rate at which a signal can be sampled (called the Nyquist rate) without such a loss of information. The theorem states that we should sample at double the maximum desired frequency so that every cycle of the original signal will be sampled at at least two points. Recall that the frequency of a sinusoidal function is the reciprocal of the period. Identify the frequency of the function f and determine the number of partitions of the interval $[0, 10]$ that give us the Nyquist rate.
- Humans cannot typically pick up signals above 20 kHz. Explain why, then, that information on a compact disk is sampled at 44,100 Hz.

Answer.

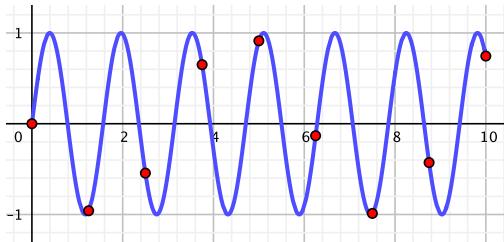
- The data points do not appear periodic at all.



- b. At least 13 samples, so at least every $10/13 \approx 0.76923$ seconds.
- c. 44100 Hz is slightly more than double 20 KHz.

Solution.

- a. Sampling the function at the endpoints of 8 equal subintervals of length $10/8 = 1.25$, we see the data pictured in the following figure (with the sample points in red). While the function is periodic with a period of $\frac{2\pi}{4} = \frac{\pi}{2}$, the 9 data points do not appear periodic at all as none of them has repeated (yet) over the course of the pictured interval.



- b. Since the period of the function $f(x) = \sin(4x)$ is $P = \frac{\pi}{2}$, its frequency is $\frac{2}{\pi} \approx 0.6366$, which means that this function completes about 0.6366 cycles per second. Doubling this frequency gives us a value of $\frac{4}{\pi} \approx 1.2732$. This means that we should sample the function at at least a rate of twice per cycle, or at least 1.2732 samples per second. On a 10-second interval, we therefore need at least 13 samples, so we should do so at least every $10/13 \approx 0.76923$ seconds.
- c. Since the theorem says to sample at a minimum of double the maximum desired frequency, and humans can't detect signals above 20 KHz, it follows that double 20 KHz is 40 KHz, or 40000 Hz, so sampling at 44100 Hz slightly exceeds the minimum sampling needed by the formal result.

8.2 Geometric Series

Motivating Questions

- What is a geometric series?
- What is a partial sum of a geometric series? What is a simplified form of the n th partial sum of a geometric series?
- Under what conditions does a geometric series converge? What is the sum of a convergent geometric series?

Many important sequences are generated by addition. In Example 8.2.1, we see an example of a sequence that is connected to a sum.

Example 8.2.1. Warfarin is an anticoagulant that prevents blood clotting; often it is prescribed to stroke victims in order to help ensure blood flow. The level of warfarin has to reach a certain concentration in the blood in order to be effective.

Suppose warfarin is taken by a particular patient in a 5 mg dose each day. The drug is absorbed by the body and some is excreted from the system between doses. Assume that at the end of a 24 hour period, 8% of the drug remains in the body. Let $Q(n)$ be the amount (in mg) of warfarin in the body before the $(n + 1)$ st dose of the drug is administered.

- Explain why $Q(1) = 5 \times 0.08$ mg.
- Explain why $Q(2) = (5 + Q(1)) \times 0.08$ mg. Then show that

$$Q(2) = (5 \times 0.08)(1 + 0.08) \text{ mg.}$$

- Explain why $Q(3) = (5 + Q(2)) \times 0.08$ mg. Then show that

$$Q(3) = (5 \times 0.08)(1 + 0.08 + 0.08^2) \text{ mg.}$$

- Explain why $Q(4) = (5 + Q(3)) \times 0.08$ mg. Then show that

$$Q(4) = (5 \times 0.08)(1 + 0.08 + 0.08^2 + 0.08^3) \text{ mg.}$$

- There is a pattern that you should see emerging. Use this pattern to find a formula for $Q(n)$, where n is an arbitrary positive integer.
- Complete Table 8.2.2 with values of $Q(n)$ for the provided n -values (reporting $Q(n)$ to 10 decimal places). What appears to be happening to the sequence $Q(n)$ as n increases?

n	1	2	3	4	5	6	7	8	9	10
$Q(n)$	0.40									

Table 8.2.2: Values of $Q(n)$ for selected values of n

8.2.1 Geometric Series

In Example 8.2.1 we encountered the sum

$$(5 \times 0.08)(1 + 0.08 + 0.08^2 + 0.08^3 + \cdots + 0.08^{n-1})$$

for the long-term level of Warfarin in the patient's system. This sum has the form

$$a + ar + ar^2 + \cdots + ar^{n-1} \quad (8.2.1)$$

where $a = 5 \times 0.08$ and $r = 0.08$. Such a sum is called a *finite geometric series* with ratio r .

Example 8.2.3. Let a and r be real numbers (with $r \neq 1$) and let

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1}.$$

In this example we will find a shortcut formula for S_n that does not involve a sum of n terms.

- Multiply S_n by r . What does the resulting sum look like?
- Subtract rS_n from S_n and explain why

$$S_n - rS_n = a - ar^n. \quad (8.2.2)$$

- Solve equation (8.2.2) for S_n to find a simple formula for S_n that does not involve adding n terms.

Hint.

- Distribute the factor of r .
- Look for common terms in the two expressions being subtracted.
- Observe that you can remove a factor of S_n from $S_n - rS_n$.

Answer.

a. $rS_n = ar + ar^2 + ar^3 + \cdots + ar^n.$

b. $S_n - rS_n = a - ar^n.$

c. $S_n = a \frac{1-r^n}{1-r}.$

Solution.

- Note that

$$rS_n = ar + ar^2 + ar^3 + \cdots + ar^n.$$

b. When we subtract rS_n from S_n the middle terms all cancel and we are left with

$$\begin{aligned} S_n - rS_n &= (a + ar + ar^2 + \cdots + ar^{n-1}) \\ &\quad - (ar + ar^2 + ar^3 + \cdots + ar^n) \\ &= a + (ar - ar) + (ar^2 - ar^2) + \cdots + (ar^{n-1} - ar^{n-1}) - ar^n \\ &= a - ar^n. \end{aligned}$$

c. Factoring S_n from left hand side and dividing gives us a formula for S_n :

$$\begin{aligned} S_n(1 - r) &= a - ar^n \\ S_n &= a \frac{1 - r^n}{1 - r}. \end{aligned}$$

The sum of the terms of a sequence is called a *series*. We summarize the result of Example 8.2.3 in the following way.

A finite geometric series S_n is a sum of the form

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1}, \quad (8.2.3)$$

where a and r are real numbers such that $r \neq 1$. The finite geometric series S_n can be written more simply as

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1} = \frac{a(1 - r^n)}{1 - r}. \quad (8.2.4)$$

We now apply Equation (8.2.4) to the example involving Warfarin from Example 8.2.1. Recall that

$$Q(n) = (5 \times 0.08)(1 + 0.08 + 0.08^2 + 0.08^3 + \cdots + 0.08^{n-1}) \text{ mg},$$

so $Q(n)$ is a geometric series with $a = 5 \times 0.08 = 0.4$ and $r = 0.08$. Thus,

$$Q(n) = 0.4 \left(\frac{1 - 0.08^n}{1 - 0.08} \right) = \frac{1}{2.3} (1 - 0.08^n).$$

Notice that as n goes to infinity, the value of 0.08^n goes to 0. So,

$$\lim_{n \rightarrow \infty} Q(n) = \lim_{n \rightarrow \infty} \frac{1}{2.3} (1 - 0.08^n) = \frac{1}{2.3} \approx 0.435.$$

Therefore, the long-term level of Warfarin in the blood under these conditions is $\frac{1}{2.3}$, which is approximately 0.435 mg.

To determine the long-term effect of Warfarin, we considered a finite geometric series of n terms, and then considered what happened as n was allowed to grow without bound. In this sense, we were actually interested in an infinite geometric series (the result of letting n go to infinity in the finite sum).

Definition 8.2.4. An infinite geometric series is an infinite sum of the form

$$a + ar + ar^2 + \cdots = \sum_{n=0}^{\infty} ar^n. \quad (8.2.5)$$

The value of r in the geometric series (8.2.5) is called the *common ratio* of the series because the ratio of the $(n + 1)$ st term, ar^n , to the n th term, ar^{n-1} , is always r :

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

Geometric series are common in mathematics and arise naturally in many different situations. As a familiar example, suppose we want to write the number with repeating decimal expansion

$$N = 0.1212\overline{12}$$

as a rational number. Observe that

$$\begin{aligned} N &= 0.12 + 0.0012 + 0.000012 + \cdots \\ &= \left(\frac{12}{100}\right) + \left(\frac{12}{100}\right)\left(\frac{1}{100}\right) + \left(\frac{12}{100}\right)\left(\frac{1}{100}\right)^2 + \cdots. \end{aligned}$$

This is an infinite geometric series with $a = \frac{12}{100}$ and $r = \frac{1}{100}$.

By using the formula for the value of a finite geometric sum, we can also develop a formula for the value of an infinite geometric series. We explore this idea in the following example.

Example 8.2.5. Let $r \neq 1$ and a be real numbers and let

$$S = a + ar + ar^2 + \cdots + ar^{n-1} + \cdots$$

be an infinite geometric series. For each positive integer n , let

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1}.$$

Recall that

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

- What should we allow n to approach in order to have S_n approach S ?
- What is the value of $\lim_{n \rightarrow \infty} r^n$ for $|r| > 1$? for $|r| < 1$? Explain.
- If $|r| < 1$, use the formula for S_n and your observations in (a) and (b) to explain why S is finite and find a resulting formula for S .

Hint.

- Let n increase without bound.
- Think about what happens to powers of numbers that are less than or greater than 1.

- c. Consider $\frac{1-r^n}{1-r}$ and how the numerator tends to 1 as $n \rightarrow \infty$ for certain values of r .

Answer.

- a. Observe that

$$S = \lim_{n \rightarrow \infty} S_n.$$

- b. If $r > 1$, then $\lim_{n \rightarrow \infty} r^n = \infty$. If $0 < r < 1$, then $\lim_{n \rightarrow \infty} r^n = 0$.

- c. Since

$$S = \lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} a \frac{1-r^n}{1-r}$$

and

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

for $0 < r < 1$, we conclude that

$$S = \lim_{n \rightarrow \infty} a \frac{1-r^n}{1-r} = \frac{a}{1-r}$$

when $0 < r < 1$.

Solution.

- a. Observe that

$$S = \lim_{n \rightarrow \infty} S_n.$$

- b. If $r > 1$, then $\lim_{n \rightarrow \infty} r^n = \infty$. If $0 < r < 1$, then $\lim_{n \rightarrow \infty} r^n = 0$.

- c. Since

$$S = \lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} a \frac{1-r^n}{1-r}$$

and

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

for $0 < r < 1$, we conclude that

$$S = \lim_{n \rightarrow \infty} a \frac{1-r^n}{1-r} = \frac{a}{1-r}$$

when $0 < r < 1$.

We can now find the value of the geometric series

$$N = \left(\frac{12}{100}\right) + \left(\frac{12}{100}\right)\left(\frac{1}{100}\right) + \left(\frac{12}{100}\right)\left(\frac{1}{100}\right)^2 + \dots$$

Using $a = \frac{12}{100}$ and $r = \frac{1}{100}$, we see that

$$N = \frac{12}{100} \left(\frac{1}{1 - \frac{1}{100}} \right) = \frac{12}{100} \left(\frac{100}{99} \right) = \frac{4}{33}.$$

The sum of a finite number of terms of an infinite geometric series is often called a *partial sum* of the series. Thus,

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1} = \sum_{k=0}^{n-1} ar^k.$$

is called the n th partial sum of the series $\sum_{k=0}^{\infty} ar^k$. We summarize our recent work with geometric series as follows.

- An infinite geometric series is an infinite sum of the form

$$a + ar + ar^2 + \cdots = \sum_{n=0}^{\infty} ar^n, \quad (8.2.6)$$

where a and r are real numbers such that $r \neq 0$.

- The n th partial sum S_n of an infinite geometric series is

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1}.$$

- If $|r| < 1$, then using the fact that $S_n = a \frac{1-r^n}{1-r}$, it follows that the sum S of the infinite geometric series (8.2.6) is

$$S = \lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} a \frac{1-r^n}{1-r} = \frac{a}{1-r}$$

Example 8.2.6. The formulas we have derived for an infinite geometric series and its partial sum have assumed we begin indexing the sums at $n = 0$. If instead we have a sum that does not begin at $n = 0$, we can factor out common terms and use the established formulas. This process is illustrated in this example.

- Consider the sum

$$\sum_{k=1}^{\infty} (2) \left(\frac{1}{3}\right)^k = (2) \left(\frac{1}{3}\right) + (2) \left(\frac{1}{3}\right)^2 + (2) \left(\frac{1}{3}\right)^3 + \cdots.$$

Remove the common factor of $(2) \left(\frac{1}{3}\right)$ from each term and hence find the sum of the series.

- Next let a and r be real numbers with $-1 < r < 1$. Consider the sum

$$\sum_{k=3}^{\infty} ar^k = ar^3 + ar^4 + ar^5 + \cdots.$$

Remove the common factor of ar^3 from each term and find the sum of the series.

- c. Finally, we consider the most general case. Let a and r be real numbers with $-1 < r < 1$, let n be a positive integer, and consider the sum

$$\sum_{k=n}^{\infty} ar^k = ar^n + ar^{n+1} + ar^{n+2} + \dots$$

Remove the common factor of ar^n from each term to find the sum of the series.

Hint.

- Think about how $r = \frac{1}{3}$.
- Note that $ar^3 + ar^4 + ar^5 + \dots = ar^3(1 + r + r^2 + \dots)$.
- Compare your work in (b).

Answer.

- $(2)\left(\frac{1}{3}\right)\left[1 + \left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)^2 + \dots\right]$.
- $ar^3 + ar^4 + ar^5 + \dots = \frac{ar^3}{1-r}$.
- $r^n + ar^{n+1} + ar^{n+2} + \dots = \frac{ar^n}{1-r}$.

Solution.

- Factoring out $(2)\left(\frac{1}{3}\right)$ gives us

$$\begin{aligned} (2)\left(\frac{1}{3}\right)\left[1 + \left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)^2 + \dots\right] &= (2)\left(\frac{1}{3}\right) \sum_{k=0}^{\infty} \left(\frac{1}{3}\right)^k \\ &= (2)\left(\frac{1}{3}\right) \left(\frac{1}{1 - \frac{1}{3}}\right) \\ &= (2)\left(\frac{1}{3}\right) \left(\frac{3}{2}\right) \\ &= 4. \end{aligned}$$

- Factoring out ar^3 gives us

$$\begin{aligned} ar^3 + ar^4 + ar^5 + \dots &= ar^3(1 + r + r^2 + \dots) \\ &= ar^3 \sum_{k=0}^{\infty} r^k \\ &= ar^3 \left(\frac{1}{1-r}\right) \\ &= \frac{ar^3}{1-r}. \end{aligned}$$

c. Factoring out ar^n gives us

$$\begin{aligned}
 ar^n + ar^{n+1} + ar^{n+2} + \cdots &= ar^n(1 + r + r^2 + \cdots) \\
 &= ar^n \sum_{k=0}^{\infty} r^k \\
 &= ar^n \left(\frac{1}{1-r} \right) \\
 &= \frac{ar^n}{1-r}.
 \end{aligned}$$

8.2.2 Summary

- An infinite geometric series is an infinite sum of the form

$$\sum_{k=0}^{\infty} ar^k$$

where a and r are real numbers and $r \neq 0$.

- The n th partial sum of the geometric series $\sum_{k=0}^{\infty} ar^k$ is

$$S_n = \sum_{k=0}^{n-1} ar^k.$$

A formula for the n th partial sum of a geometric series is

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

If $|r| < 1$, the infinite geometric series $\sum_{k=0}^{\infty} ar^k$ has the finite sum $\frac{a}{1-r}$.

8.2.3 Exercises

1. (Seventh term of a geometric sequence) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (A geometric series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (A series that is not geometric) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Two sums of geometric sequences) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. There is an old question that is often used to introduce the power of geometric growth. Here is one version. Suppose you are hired for a one month (30 days, working every day) job and are given two options to be paid.

Option 1. You can be paid \$500 per day or

Option 2. You can be paid 1 cent the first day, 2 cents the second day, 4 cents the third day, 8 cents the fourth day, and so on, doubling the amount you are paid each day.

- How much will you be paid for the job in total under Option 1?
- Complete Table 8.2.7 to determine the pay you will receive under Option 2 for the first 10 days.

Day	Pay on this day	Total amount paid to date
1	\$0.01	\$0.01
2	\$0.02	\$0.03
3		
4		
5		
6		
7		
8		
9		
10		

Table 8.2.7: Option 2 payments

- Find a formula for the amount paid on day n , as well as for the total amount paid by day n . Use this formula to determine which option (1 or 2) you should take.

Answer.

- $30 \cdot 500 = 1500$ dollars.
-

Day	Pay on this day	Total amount paid to date
1	\$0.01	\$0.01
2	\$0.02	\$0.03
3	\$0.04	\$0.07
4	\$0.08	\$0.15
5	\$0.16	\$0.31
6	\$0.32	\$0.63
7	\$0.64	\$1.27
8	\$1.28	\$2.55
9	\$2.56	\$5.11
10	\$5.12	\$10.23

- $\$0.01(2^{30} - 1) = \$10,737,418.23$.

Solution.

- You will be paid $30 \cdot 500 = 1500$ dollars under Option 1.

b.

Day	Pay on this day	Total amount paid to date
1	\$0.01	\$0.01
2	\$0.02	\$0.03
3	\$0.04	\$0.07
4	\$0.08	\$0.15
5	\$0.16	\$0.31
6	\$0.32	\$0.63
7	\$0.64	\$1.27
8	\$1.28	\$2.55
9	\$2.56	\$5.11
10	\$5.12	\$10.23

- c. On day n you are paid $0.01(2^{n-1})$ dollars and the total amount paid is $0.01(2^n - 1)$.
So on day 30 under Option 2 the total payment will be

$$0.01(2^{30} - 1) = \$10,737,418.23,$$

so Option 2 is definitely the better deal.

6. Suppose you drop a golf ball onto a hard surface from a height h . The collision with the ground causes the ball to lose energy and so it will not bounce back to its original height. The ball will then fall again to the ground, bounce back up, and continue. Assume that at each bounce the ball rises back to a height $\frac{3}{4}$ of the height from which it dropped. Let h_n be the height of the ball on the n th bounce, with $h_0 = h$. In this exercise we will determine the distance traveled by the ball and the time it takes to travel that distance.

- Determine a formula for h_1 in terms of h .
- Determine a formula for h_2 in terms of h .
- Determine a formula for h_3 in terms of h .
- Determine a formula for h_n in terms of h .
- Write an infinite series that represents the total distance traveled by the ball. Then determine the sum of this series.
- Next, let's determine the total amount of time the ball is in the air.
 - When the ball is dropped from a height H , if we assume the only force acting on it is the acceleration due to gravity, then the height of the ball at time t is given by

$$H - \frac{1}{2}gt^2.$$

Use this formula to determine the time it takes for the ball to hit the ground after being dropped from height H .

- Use your work in the preceding item, along with that in (a)-(e) above to determine the total amount of time the ball is in the air.

Answer.

a. $h_1 = \left(\frac{3}{4}\right)h$.

- b. $h_2 = \left(\frac{3}{4}\right)h_1 = \left(\frac{3}{4}\right)^2 h.$
- c. $h_3 = \left(\frac{3}{4}\right)h_2 = \left(\frac{3}{4}\right)^3 h.$
- d. $h_n = \left(\frac{3}{4}\right)h_{n-1} = \left(\frac{3}{4}\right)^n h.$
- e. The distance traveled by the ball is $7h$, which is finite.

Solution.

- a. Since the ball bounces back $3/4$ of the original height, we have

$$h_1 = \left(\frac{3}{4}\right)h.$$

- b. Since the ball bounces back $3/4$ of the original height, we have

$$h_2 = \left(\frac{3}{4}\right)h_1 = \left(\frac{3}{4}\right)^2 h.$$

- c. Since the ball bounces back $3/4$ of the original height, we have

$$h_3 = \left(\frac{3}{4}\right)h_2 = \left(\frac{3}{4}\right)^3 h.$$

- d. Since the ball bounces back $3/4$ of the original height, we have

$$h_n = \left(\frac{3}{4}\right)h_{n-1} = \left(\frac{3}{4}\right)^n h.$$

- e. The distance traveled by the ball is

$$\begin{aligned} h + 2[h_1 + h_2 + \dots] &= h + 2 \left[\left(\frac{3}{4}\right)^2 h + \left(\frac{3}{4}\right)^2 h + \left(\frac{3}{4}\right)^2 h + \dots \right] \\ &= h + 2 \left(\frac{3}{4}\right) h \left[1 + \left(\frac{3}{4}\right) + \left(\frac{3}{4}\right)^2 + \dots \right] \\ &= h + \left(\frac{3}{2}\right) h \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \\ &= h + \left(\frac{3}{2}\right) h \left(\frac{1}{1 - 3/4}\right) \\ &= h + \left(\frac{3}{2}\right) h(4) \\ &= h + 6h \\ &= 7h. \end{aligned}$$

So no matter what the initial height of the ball, the ball will bounce infinitely many times but only travel a finite distance.

7. Suppose you play a game with a friend that involves rolling a standard six-sided die. Before a player can participate in the game, he or she must roll a six with the die. Assume that you roll first and that you and your friend take alternate rolls. In this exercise we will determine the probability that you roll the first six.

- Explain why the probability of rolling a six on any single roll (including your first turn) is $\frac{1}{6}$.
- If you don't roll a six on your first turn, then in order for you to roll the first six on your second turn, both you and your friend had to fail to roll a six on your first turns, and then you had to succeed in rolling a six on your second turn. Explain why the probability of this event is

$$\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right).$$

- Now suppose you fail to roll the first six on your second turn. Explain why the probability is

$$\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)$$

that you to roll the first six on your third turn.

- The probability of you rolling the first six is the probability that you roll the first six on your first turn plus the probability that you roll the first six on your second turn plus the probability that you roll the first six on your third turn, and so on. Explain why this probability is

$$\frac{1}{6} + \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) + \dots$$

Find the sum of this series and determine the probability that you roll the first six.

Answer.

- There are 6 equally possible outcomes when we roll one die.
- The three rolls are independent so the probability of the overall outcome is the product of the three probabilities.
- See (b).
- $P = \frac{6}{11}$.

Solution.

- There are 6 equally possible outcomes when we roll one die, and only one of those is a 6, so the probability that a 6 appears when we roll the die is $\frac{1}{6}$.
- On the first roll, the probability of not rolling a 6 is $\frac{5}{6}$. This is the same on the second roll, so the probability of not rolling a 6 on either of the first two rolls is $\frac{5}{6} \cdot \frac{5}{6}$, since the two events are independent. If we then roll a 6 on the following turn, the probability of doing so in that roll is $\frac{1}{6}$, so with the three rolls being independent, the probability of you rolling a 6 on your second turn (the third overall roll) is

$$P = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right).$$

- The exact same reasoning as in (b), but with 5 dice rolls instead of 3, shows that

$$P = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right).$$

- d. By simply summing the probabilities to consider all of the possible outcomes (rolling 6 on your first turn, doing so on your second, etc.), we get that the total probability of you rolling a 6 on one of your turns is

$$P = \frac{1}{6} + \left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^4 \left(\frac{1}{6}\right) + \dots,$$

which is a geometric series with $a = \frac{1}{6}$ and $r = \left(\frac{5}{6}\right)^2$. Since we know that a geometric series has sum $S = \frac{a}{1-r}$, it follows that

$$P = \frac{\frac{1}{6}}{1 - \frac{25}{36}} = \frac{\frac{6}{36}}{\frac{11}{36}} = \frac{6}{11},$$

so by going first your odds are just slightly better than half that you'll be the first one to roll a 6.

8. The goal of a federal government stimulus package is to positively affect the economy. Economists and politicians quote numbers like “ k million jobs and a net stimulus to the economy of n billion of dollars.” Where do they get these numbers? Let’s consider one aspect of a stimulus package: tax cuts. Economists understand that tax cuts or rebates can result in long-term spending that is many times the amount of the rebate. For example, assume that for a typical person, 75% of her entire income is spent (that is, put back into the economy). Further, assume the government provides a tax cut or rebate that totals P dollars for each person.

- The tax cut of P dollars is income for its recipient. How much of this tax cut will be spent?
- In this simple model, we will say that the spent portion of the tax cut/rebate from part (a) then becomes income for another person who, in turn, spends 75% of this income. After this “second round” of spent income, how many total dollars have been added to the economy as a result of the original tax cut/rebate?
- This second round of spending becomes income for another group who spend 75% of this income, and so on. In economics this is called the *multiplier effect*. Explain why an original tax cut/rebate of P dollars will result in multiplied spending of

$$0.75P(1 + 0.75 + 0.75^2 + \dots).$$

dollars.

- Based on these assumptions, how much stimulus will a 200 billion dollar tax cut/rebate to consumers add to the economy, assuming consumer spending remains consistent forever.

Answer.

- $0.75P$ dollars spent.
- $0.75P + 0.75(0.75P) = 0.75P(1 + 0.75)$ dollars.
- $0.75P + 0.75^2P + 0.75^2P + \dots = 0.75P(1 + 0.75 + 0.75^2 + \dots)$ dollars.
- A stimulus of 200 billion dollars adds 600 billion dollars to the economy.

Solution.

- Given that 75% of income is spent, this tax cut/rebate will result in $0.75P$ dollars spent.
- The amount of money introduced into the economy is the amount $0.75P$ after the original cut plus 75% of that income from spending. So the total amount of money added to the economy is

$$0.75P + 0.75(0.75P) = 0.75P(1 + 0.75)$$

dollars.

- The $n - 1$ st round of spending will result in income of $0.75^n P$ dollars, of which 75% or $0.75^{n+1}P$ will be spent. So each round of spending adds a factor to the total amount, giving us a total amount of spending in the economy of

$$0.75P + 0.75^2P + 0.75^3P + \dots = 0.75P(1 + 0.75 + 0.75^2 + \dots)$$

dollars.

- Using the sum of a geometric series with ratio 0.75 shows that

$$\begin{aligned} 0.75P(1 + 0.75 + 0.75^2 + \dots) &= 0.75P \left(\frac{1}{1 - 0.75} \right) \\ &= 0.75P(4) \\ &= 3P. \end{aligned}$$

So a stimulus of 200 billion dollars adds 600 billion dollars to the economy.

- Like stimulus packages, home mortgages and foreclosures also impact the economy. A problem for many borrowers is the adjustable rate mortgage, in which the interest rate can change (and usually increases) over the duration of the loan, causing the monthly payments to increase beyond the ability of the borrower to pay. Most financial analysts recommend fixed rate loans, ones for which the monthly payments remain constant throughout the term of the loan. In this exercise we will analyze fixed rate loans.

When most people buy a large ticket item like car or a house, they have to take out a loan to make the purchase. The loan is paid back in monthly installments until the entire amount of the loan, plus interest, is paid. With a loan, we borrow money, say P dollars (called the principal), and pay off the loan at an interest rate of $r\%$. To pay back the loan we make regular monthly payments, some of which goes to pay off the principal and some of which is charged as interest. In most cases, the interest is computed based on the amount of principal that remains at the beginning of the month. We assume a fixed rate loan, that is one in which we make a constant monthly payment M on our loan, beginning in the original month of the loan.

Suppose you want to buy a house. You have a certain amount of money saved to make a down payment, and you will borrow the rest to pay for the house. Of course, for the privilege of loaning you the money, the bank will charge you interest on this loan, so the amount you pay back to the bank is more than the amount you borrow. In fact, the amount you ultimately pay depends on three things: the amount you borrow (called the *principal*), the interest rate, and the length of time you have to pay off the loan plus interest (called the *duration* of the loan). For this example, we assume that the interest rate is fixed at $r\%$.

To pay off the loan, each month you make a payment of the same amount (called *installments*). Suppose we borrow P dollars (our principal) and pay off the loan at an interest rate of $r\%$

with regular monthly installment payments of M dollars. So in month 1 of the loan, before we make any payments, our principal is P dollars. Our goal in this exercise is to find a formula that relates these three parameters to the time duration of the loan.

We are charged interest every month at an annual rate of $r\%$, so each month we pay $\frac{r}{12}\%$ interest on the principal that remains. Given that the original principal is P dollars, we will pay $(\frac{0.0r}{12})P$ dollars in interest on our first payment. Since we paid M dollars in total for our first payment, the remainder of the payment $(M - (\frac{r}{12})P)$ goes to pay down the principal. So the principal remaining after the first payment (let's call it P_1) is the original principal minus what we paid on the principal, or

$$P_1 = P - \left(M - \left(\frac{r}{12} \right) P \right) = \left(1 + \frac{r}{12} \right) P - M.$$

As long as P_1 is positive, we still have to keep making payments to pay off the loan.

- Recall that the amount of interest we pay each time depends on the principal that remains. How much interest, in terms of P_1 and r , do we pay in the second installment?
- How much of our second monthly installment goes to pay off the principal? What is the principal P_2 , or the balance of the loan, that we still have to pay off after making the second installment of the loan? Write your response in the form $P_2 = ()P_1 - ()M$, where you fill in the parentheses.
- Show that $P_2 = (1 + \frac{r}{12})^2 P - [1 + (1 + \frac{r}{12})] M$.
- Let P_3 be the amount of principal that remains after the third installment. Show that

$$P_3 = \left(1 + \frac{r}{12} \right)^3 P - \left[1 + \left(1 + \frac{r}{12} \right) + \left(1 + \frac{r}{12} \right)^2 \right] M.$$

- If we continue in the manner described in the problems above, then the remaining principal of our loan after n installments is

$$P_n = \left(1 + \frac{r}{12} \right)^n P - \left[\sum_{k=0}^{n-1} \left(1 + \frac{r}{12} \right)^k \right] M. \quad (8.2.7)$$

This is a rather complicated formula and one that is difficult to use. However, we can simplify the sum if we recognize part of it as a partial sum of a geometric series. Find a formula for the sum

$$\sum_{k=0}^{n-1} \left(1 + \frac{r}{12} \right)^k. \quad (8.2.8)$$

and then a general formula for P_n that does not involve a sum.

- It is usually more convenient to write our formula for P_n in terms of years rather than months. Show that $P(t)$, the principal remaining after t years, can be written as

$$P(t) = \left(P - \frac{12M}{r} \right) \left(1 + \frac{r}{12} \right)^{12t} + \frac{12M}{r}. \quad (8.2.9)$$

- Now that we have analyzed the general loan situation, we apply formula (8.2.9) to an actual loan. Suppose we charge \$1,000 on a credit card for holiday expenses. If our

credit card charges 20% interest and we pay only the minimum payment of \$25 each month, how long will it take us to pay off the \$1,000 charge? How much in total will we have paid on this \$1,000 charge? How much total interest will we pay on this loan?

- h. Now we consider larger loans, e.g., automobile loans or mortgages, in which we borrow a specified amount of money over a specified period of time. In this situation, we need to determine the amount of the monthly payment we need to make to pay off the loan in the specified amount of time. In this situation, we need to find the monthly payment M that will take our outstanding principal to 0 in the specified amount of time. To do so, we want to know the value of M that makes $P(t) = 0$ in formula (8.2.9). If we set $P(t) = 0$ and solve for M , it follows that

$$M = \frac{rP(1 + \frac{r}{12})^{12t}}{12((1 + \frac{r}{12})^{12t} - 1)}.$$

- i) Suppose we want to borrow \$15,000 to buy a car. We take out a 5 year loan at 6.25%. What will our monthly payments be? How much in total will we have paid for this \$15,000 car? How much total interest will we pay on this loan?
- ii) Suppose you charge your books for winter semester on your credit card. The total charge comes to \$525. If your credit card has an interest rate of 18% and you pay \$20 per month on the card, how long will it take before you pay off this debt? How much total interest will you pay?
- iii) Say you need to borrow \$100,000 to buy a house. You have several options on the loan:
 - 30 years at 6.5%
 - 25 years at 7.5%
 - 15 years at 8.25%.
 a. What are the monthly payments for each loan?
 b. Which mortgage is ultimately the best deal (assuming you can afford the monthly payments)? In other words, for which loan do you pay the least amount of total interest?

Answer.

- a. $(\frac{r}{12})P_1$ dollars.
- b. $P_2 = (1 + \frac{r}{12})P_1 - M$.
- c. $P_2 = (1 + \frac{r}{12})^2 P - [1 + (1 + \frac{r}{12})] M$.
- d. $P_3 = (1 + \frac{r}{12})P_2 - M$.
- $$P_3 = (1 + \frac{r}{12})^3 P - \left[1 + (1 + \frac{r}{12}) + (1 + \frac{r}{12})^2\right] M.$$
- e.
$$P_n = P(1 + \frac{r}{12})^n - \left(\frac{12M}{r}\right) \left((1 + \frac{r}{12})^n - 1\right).$$
- f.
$$P(t) = P(1 + \frac{r}{12})^{12t} - \left(\frac{12M}{r}\right) \left((1 + \frac{r}{12})^{12t} - 1\right)$$
- g.
$$A(t) = \left(1000 - \frac{12(25)}{0.2}\right) (1 + \frac{0.2}{12})^{12t} + \frac{12(25)}{0.2}$$

$$t \approx 5.5.$$

We pay \$659 dollars in interest on our \$1000 loan.

- h. \$291.74 each month to complete the loan in 5 years; we pay \$2,504.40 in interest.

Solution.

- a. In the second installment, we are charged $(\frac{r}{12})P_1$ dollars in interest.
 b. The amount of our payment M that doesn't go toward interest is applied toward the principal. So we reduce our principal by $M - (\frac{r}{12})P_1$ dollars. Therefore, the balance of our loan after the second installment is

$$P_2 = P_1 - \left(M - \left(\frac{r}{12} \right) P_1 \right) = \left(1 + \frac{r}{12} \right) P_1 - M.$$

- c. We know $P_2 = (1 + \frac{r}{12})P_1 - M$, so we substitute for P_1 and simplify:

$$\begin{aligned} P_2 &= \left(1 + \frac{r}{12} \right) P_1 - M \\ &= \left(1 + \frac{r}{12} \right) \left[\left(1 + \frac{r}{12} \right) P - M \right] - M \\ &= \left(1 + \frac{r}{12} \right)^2 P - \left[1 + \left(1 + \frac{r}{12} \right) \right] M. \end{aligned}$$

- d. For the third payment we are charged $(\frac{r}{12})P_2$ dollars in interest. So $M - (\frac{r}{12})P_2$ dollars goes toward paying down the principal. Then

$$P_3 = P_2 - \left(M - \left(\frac{r}{12} \right) P_2 \right) = \left(1 + \frac{r}{12} \right) P_2 - M.$$

- So $P_3 = (1 + \frac{r}{12})P_2 - M$, so we substitute for P_2 and simplify:

$$\begin{aligned} P_3 &= \left(1 + \frac{r}{12} \right) P_2 - M \\ &= \left(1 + \frac{r}{12} \right) \left[\left(1 + \frac{r}{12} \right)^2 P - \left[1 + \left(1 + \frac{r}{12} \right) \right] M \right] - M \\ &= \left(1 + \frac{r}{12} \right)^3 P - \left[1 + \left(1 + \frac{r}{12} \right) + \left(1 + \frac{r}{12} \right)^2 \right] M. \end{aligned}$$

- e. The sum in (8.2.8) is a partial sum of the geometric series with ratio $1 + \frac{r}{12}$. So formula (8.2.4) shows that the partial sum $\sum_{k=0}^{n-1} \left(1 + \frac{r}{12} \right)^k$ is equal to

$$\begin{aligned} \sum_{k=0}^{n-1} \left(1 + \frac{r}{12} \right)^k &= \frac{1 - \left(1 + \frac{r}{12} \right)^n}{1 - \left(1 + \frac{r}{12} \right)} \\ &= \left(\frac{12}{r} \right) \left(\left(1 + \frac{r}{12} \right)^n - 1 \right). \end{aligned}$$

Now we can determine that the remaining principal after n installments is

$$\begin{aligned} P_n &= \left(1 + \frac{r}{12} \right)^n P - \left[\left(\frac{12}{r} \right) \left(\left(1 + \frac{r}{12} \right)^n - 1 \right) \right] M \\ &= P \left(1 + \frac{r}{12} \right)^n - \left(\frac{12M}{r} \right) \left(\left(1 + \frac{r}{12} \right)^n - 1 \right). \end{aligned}$$

- f. To put our payments into a yearly context which is more convenient, we know that there are $12t$ months in t years. So we can say that the principal $P(t)$ remaining at the end of t years is

$$P(t) = P \left(1 + \frac{r}{12}\right)^{12t} - \left(\frac{12M}{r}\right) \left(\left(1 + \frac{r}{12}\right)^{12t} - 1\right) \quad (8.2.10)$$

which reduces to (8.2.9).

- g. Under these conditions, the principal in the account at the end of t years will be

$$A(t) = \left(1000 - \frac{12(25)}{0.2}\right) \left(1 + \frac{0.2}{12}\right)^{12t} + \frac{12(25)}{0.2}.$$

To determine how long it will take us to pay off the card, we need to find the time so that $A(t) = 0$. When $A(t) = 0$ we have

$$\begin{aligned} 0 &= \left(1000 - \frac{12(25)}{0.2}\right) \left(1 + \frac{0.2}{12}\right)^{12t} + \frac{12(25)}{0.2} \\ 0 &= (-500)(1.0167)^{12t} + 1500 \\ 500(1.0167)^{12t} &= 1500 \\ (1.0167)^{12t} &= 3 \\ 12t \ln(1.0167) &= \ln(3) \\ t &= \frac{\ln(3)}{12 \ln(1.0167)} \\ t &\approx 5.53. \end{aligned}$$

So it would take over 5.5 years to pay off the debt. At \$25 per month, we will have paid approximately $12 \times 5.5 \times 25 = 1,659$ dollars on our original \$1,000 charge. In other words, we pay \$659 dollars in interest on our \$1000 loan.

- h. We need to find the value of M in our formula:

$$M = \frac{rP \left(1 + \frac{r}{12}\right)^{12t}}{12 \left(\left(1 + \frac{r}{12}\right)^{12t} - 1\right)} = \frac{(0.0625)(15000) \left(1 + \frac{0.0625}{12}\right)^{60}}{12 \left(\left(1 + \frac{0.0625}{12}\right)^{60} - 1\right)} \approx 291.74.$$

So we need to pay \$291.74 each month to complete the loan in 5 years. During that time we spend a total of $12 \times 5 \times 291.74 = 17,504.40$ dollars on the loan or \$2,504.40 in interest.

8.3 Series of Real Numbers

Motivating Questions

- What is an infinite series?
- What is the n th partial sum of an infinite series?
- How do we add up an infinite number of numbers? In other words, what does it mean for an infinite series of real numbers to converge?
- What does it mean for an infinite series of real numbers to diverge?

In Section 8.2, we encountered infinite geometric series. For example, by writing

$$N = 0.1212121212 \cdots = \frac{12}{100} + \frac{12}{100} \cdot \frac{1}{100} + \frac{12}{100} \cdot \frac{1}{100^2} + \cdots$$

as a geometric series, we found a way to write N as a single fraction: $N = \frac{4}{33}$. In this section, we explore other types of infinite sums. In Example 8.3.1 we see how one such sum is related to the famous number e .

Example 8.3.1. Have you ever wondered how your calculator can produce a numeric approximation for complicated numbers like e , π or $\ln(2)$? After all, the only operations a calculator can really perform are addition, subtraction, multiplication, and division, the operations that make up polynomials. This example provides the first steps in understanding how this process works. Throughout the example, let $f(x) = e^x$.

- Find the tangent line to f at $x = 0$ and use this linearization to approximate e . That is, find a formula $L(x)$ for the tangent line, and compute $L(1)$, since $L(1) \approx f(1) = e$.
- The linearization of e^x does not provide a good approximation to e since 1 is not very close to 0. To obtain a better approximation, we alter our approach a bit. Instead of using a straight line to approximate e , we put an appropriate bend in our estimating function to make it better fit the graph of e^x for x close to 0. With the linearization, we had both $f(x)$ and $f'(x)$ share the same value as the linearization at $x = 0$. We will now use a quadratic approximation $P_2(x)$ to $f(x) = e^x$ centered at $x = 0$ which has the property that $P_2(0) = f(0)$, $P'_2(0) = f'(0)$, and $P''_2(0) = f''(0)$.
 - Let $P_2(x) = 1 + x + \frac{x^2}{2}$. Show that $P_2(0) = f(0)$, $P'_2(0) = f'(0)$, and $P''_2(0) = f''(0)$. Then, use $P_2(x)$ to approximate e by observing that $P_2(1) \approx f(1)$.
 - We can continue approximating e with polynomials of larger degree whose higher derivatives agree with those of f at 0. This turns out to make the polynomials fit the graph of f better for more values of x around 0. For example, let $P_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$. Show that $P_3(0) = f(0)$, $P'_3(0) = f'(0)$, $P''_3(0) = f''(0)$, and $P'''_3(0) = f'''(0)$. Use $P_3(x)$ to approximate e in a way similar to how you did so with $P_2(x)$ above.

Solution.

- a. The linearization of f at $x = a$ is

$$f(a) + f'(a)(x - a),$$

so the linearization $P_1(x)$ of $f(x) = e^x$ at $x = 0$ is

$$P_1(x) = e^0 + e^0(x - 0) = 1 + x.$$

Now

$$f(x) \approx P_1(x)$$

for x close to 0 and so

$$e = e^1 \approx P_1(1) = 1 + 1 = 2.$$

- b. i) The derivatives of P_2 and f are

$$\begin{array}{ll} P_2(x) = 1 + x + \frac{x^2}{2} & f(x) = e^x \\ P'_2(x) = 1 + x & f'(x) = e^x \\ P''_2(x) = 1 & f''(x) = e^x, \end{array}$$

and so the derivatives of P_2 and f evaluated at 0 are

$$\begin{array}{ll} P_2(0) = 1 & f(0) = e^0 = 1 \\ P'_2(0) = 1 & f'(0) = e^0 = 1 \\ P''_2(0) = 1 & f''(0) = e^0 = 1. \end{array}$$

Then

$$e = e^1 \approx P_2(1) = 1 + 1 + \frac{1}{2} = 2.5.$$

- ii) The derivatives of P_3 and f are

$$\begin{array}{ll} P_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} & f(x) = e^x \\ P'_3(x) = 1 + x + \frac{x^2}{2} & f'(x) = e^x \\ P''_3(x) = 1 + x & f''(x) = e^x \\ P'''_3(x) = 1 & f'''(x) = e^x, \end{array}$$

and so the derivatives of P_3 and f evaluated at 0 are

$$\begin{array}{ll} P_3(0) = 1 & f(0) = e^0 = 1 \\ P'_3(0) = 1 & f'(0) = e^0 = 1 \\ P''_3(0) = 1 & f''(0) = e^0 = 1 \\ P'''_3(0) = 1 & f'''(0) = e^0 = 1. \end{array}$$

Then

$$e = e^1 \approx P_3(1) = 1 + 1 + \frac{1}{2} + \frac{1}{6} = \frac{8}{3} \approx 2.67.$$

8.3.1 Infinite Series

Example 8.3.1 shows that an approximation to e using a linear polynomial is 2, an approximation to e using a quadratic polynomial is 2.5, and an approximation using a cubic polynomial is 2.6667. If we continue this process we can obtain approximations from quartic (degree 4) and quintic (degree 5) polynomials, giving us the following approximations to e :

linear	$1 + 1$	2
quadratic	$1 + 1 + \frac{1}{2}$	2.5
cubic	$1 + 1 + \frac{1}{2} + \frac{1}{6}$	2.6
quartic	$1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}$	2.7083
quintic	$1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120}$	2.716

We see an interesting pattern here. The number e is being approximated by the sum

$$1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots + \frac{1}{n!} \quad (8.3.1)$$

for increasing values of n .

We can use summation notation as a shorthand¹ for writing the sum in Equation (8.3.1) to get

$$e \approx 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots + \frac{1}{n!} = \sum_{k=0}^n \frac{1}{k!}. \quad (8.3.2)$$

We can calculate this sum using as large an n as we want, and the larger n is the more accurate the approximation (8.3.2) is. This suggests that we can write the number e as the infinite sum

$$e = \sum_{k=0}^{\infty} \frac{1}{k!}. \quad (8.3.3)$$

This sum is an example of an *infinite series*. Note that the series (8.3.3) is the sum of the terms of the infinite sequence $\{\frac{1}{n!}\}$. In general, we use the following notation and terminology.

Definition 8.3.2. An **infinite series** of real numbers is the sum of the entries in an infinite sequence of real numbers. In other words, an infinite series is sum of the form

$$a_1 + a_2 + \cdots + a_n + \cdots = \sum_{k=1}^{\infty} a_k,$$

where a_1, a_2, \dots , are real numbers.

We use summation notation to identify a series. If the series adds the entries of a sequence $\{a_n\}_{n \geq 1}$, we will write the series as

$$\sum_{k \geq 1} a_k$$

¹Note that $0!$ appears in Equation (8.3.2). By definition, $0! = 1$.

or

$$\sum_{k=1}^{\infty} a_k.$$

Each of these notations is simply shorthand for the infinite sum $a_1 + a_2 + \cdots + a_n + \cdots$.

Is it even possible to sum an infinite list of numbers? That doing so is possible in some situations shouldn't come as a surprise. The definite integral is the sum of an infinite collections of numbers, and we have already examined the special case of geometric series in the previous section. As we investigate other infinite sums, there are two key questions we seek to answer: (1) is the sum finite? and (2) if yes, what is its value?

Example 8.3.3. Consider the series

$$\sum_{k=1}^{\infty} \frac{1}{k^2}.$$

While it is physically impossible to add an infinite collection of numbers, we can, of course, add any finite collection of them. In what follows, we investigate how understanding how to find the n th partial sum (that is, the sum of the first n terms) enables us to make sense of the infinite sum.

- a. Sum the first two numbers in this series. That is, find a numeric value for

$$\sum_{k=1}^2 \frac{1}{k^2}$$

- b. Next, add the first three numbers in the series.

- c. Continue adding terms in this series to complete the list below. Carry each sum to at least 8 decimal places.

$$\sum_{k=1}^1 \frac{1}{k^2} = 1$$

$$\sum_{k=1}^6 \frac{1}{k^2} =$$

$$\sum_{k=1}^2 \frac{1}{k^2} =$$

$$\sum_{k=1}^7 \frac{1}{k^2} =$$

$$\sum_{k=1}^3 \frac{1}{k^2} =$$

$$\sum_{k=1}^8 \frac{1}{k^2} =$$

$$\sum_{k=1}^4 \frac{1}{k^2} =$$

$$\sum_{k=1}^9 \frac{1}{k^2} =$$

$$\sum_{k=1}^5 \frac{1}{k^2} =$$

$$\sum_{k=1}^{10} \frac{1}{k^2} =$$

- d. The sums in the table in part c form a sequence whose n th term is $S_n = \sum_{k=1}^n \frac{1}{k^2}$. Based on your calculations in the table, do you think the sequence $\{S_n\}$ converges or diverges? Explain. How do you think this sequence $\{S_n\}$ is related to the series $\sum_{k=1}^{\infty} \frac{1}{k^2}$?

Hint.

- Compute $\frac{1}{1^2} + \frac{1}{2^2}$.
- Do likewise but include the third term, $\frac{1}{3^2}$.
- Note that you can use your work in each previous stage and just add one more term to get the next desired sum.
- Look for a pattern in the numbers you computed in part (c).

Answer.

a. See the table in part c.

b. See the table in part c.

c.
$$\sum_{k=1}^1 \frac{1}{k^2} = 1$$

$$\sum_{k=1}^2 \frac{1}{k^2} = 1.25$$

$$\sum_{k=1}^3 \frac{1}{k^2} = 1.361111111$$

$$\sum_{k=1}^4 \frac{1}{k^2} = 1.423611111$$

$$\sum_{k=1}^5 \frac{1}{k^2} = 1.463611111$$

$$\sum_{k=1}^6 \frac{1}{k^2} = 1.491388889$$

$$\sum_{k=1}^7 \frac{1}{k^2} = 1.511797052$$

$$\sum_{k=1}^8 \frac{1}{k^2} = 1.527422052$$

$$\sum_{k=1}^9 \frac{1}{k^2} = 1.539767731$$

$$\sum_{k=1}^{10} \frac{1}{k^2} = 1.549767731$$

d. It appears $\{S_n\}$ converges to something a bit larger than 1.5.

Solution.

a. See the table in part c.

b. See the table in part c.

c. If we add the first few terms of the sequence $\left\{\frac{1}{k^2}\right\}$ we obtain the entries (to 10 decimal places) in the following table.

$$\sum_{k=1}^1 \frac{1}{k^2} = 1$$

$$\sum_{k=1}^2 \frac{1}{k^2} = 1.25$$

$$\sum_{k=1}^3 \frac{1}{k^2} = 1.361111111$$

$$\sum_{k=1}^4 \frac{1}{k^2} = 1.423611111$$

$$\sum_{k=1}^5 \frac{1}{k^2} = 1.463611111$$

$$\sum_{k=1}^6 \frac{1}{k^2} = 1.491388889$$

$$\sum_{k=1}^7 \frac{1}{k^2} = 1.511797052$$

$$\sum_{k=1}^8 \frac{1}{k^2} = 1.527422052$$

$$\sum_{k=1}^9 \frac{1}{k^2} = 1.539767731$$

$$\sum_{k=1}^{10} \frac{1}{k^2} = 1.549767731$$

- d. These sums in the table in part (c) seem to indicate that the sequence $\{S_n\}$ converges to something a bit larger than 1.5. Since the sequence $\{S_n\}$ is found by adding up the first k terms of the series, we should expect that the series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is the limit of the sequence $\{S_n\}$ as n goes to infinity.

The example in Example 8.3.3 illustrates how we define the sum of an infinite series. We construct a new sequence of numbers (called the *sequence of partial sums*) where the n th term in the sequence consists of the sum of the first n terms of the series. If this sequence converges, the corresponding infinite series is said to converge, and we say that we can find the sum of the series. More formally, we have the following definition.

Definition 8.3.4. The n th **partial sum** of the series $\sum_{k=1}^{\infty} a_k$ is the finite sum $S_n = \sum_{k=1}^n a_k$.

In other words, the n th partial sum S_n of a series is the sum of the first n terms in the series,

$$S_n = a_1 + a_2 + \cdots + a_n.$$

We investigate the behavior of the series by examining the sequence

$$S_1, S_2, \dots, S_n, \dots$$

of its partial sums. If the sequence of partial sums converges to some finite number, we say that the corresponding series *converges*. Otherwise, we say the series *diverges*. From our work in Example 8.3.3, the series

$$\sum_{k=1}^{\infty} \frac{1}{k^2}$$

appears to converge to some number near 1.54977. We formalize the concept of convergence and divergence of an infinite series in the following definition.

Definition 8.3.5. The infinite series

$$\sum_{k=1}^{\infty} a_k$$

converges (or is **convergent**) if the sequence $\{S_n\}$ of partial sums converges, where

$$S_n = \sum_{k=1}^n a_k.$$

If $\lim_{n \rightarrow \infty} S_n = S$, then we call S the sum of the series $\sum_{k=1}^{\infty} a_k$. That is,

$$\sum_{k=1}^{\infty} a_k = \lim_{n \rightarrow \infty} S_n = S.$$

If the sequence of partial sums does not converge, then the series

$$\sum_{k=1}^{\infty} a_k$$

diverges (or is divergent).

The early terms in a series do not influence whether or not the series converges or diverges. Rather, the convergence or divergence of a series

$$\sum_{k=1}^{\infty} a_k$$

is determined by what happens to the terms a_k for very large values of k . To see why, suppose that m is some constant larger than 1. Then

$$\sum_{k=1}^{\infty} a_k = (a_1 + a_2 + \cdots + a_m) + \sum_{k=m+1}^{\infty} a_k.$$

Since $a_1 + a_2 + \cdots + a_m$ is a finite number, the series $\sum_{k=1}^{\infty} a_k$ will converge if and only if the series $\sum_{k=m+1}^{\infty} a_k$ converges. Because the starting index of the series doesn't affect whether the series converges or diverges, we will often just write

$$\sum a_k$$

when we are interested in questions of convergence/divergence as opposed to the exact sum of a series.

In Section 8.2 we encountered the special family of infinite geometric series. Recall that a geometric series has the form $\sum_{k=0}^{\infty} ar^k$, where a and r are real numbers (and $r \neq 1$). We found that the n th partial sum S_n of a geometric series is given by the convenient formula

$$S_n = \frac{1 - r^n}{1 - r},$$

and thus a geometric series converges if $|r| < 1$. Geometric series diverge for all other values of r .

It is generally a difficult question to determine if a given nongeometric series converges or diverges. There are several tests we can use that we will consider in the following sections.

8.3.2 The Divergence Test

The first question we ask about any infinite series is usually “Does the series converge or diverge?” There is a straightforward way to check that certain series diverge, and we explore this test in the next example.

Example 8.3.6. If the series $\sum a_k$ converges, then an important result necessarily follows regarding the sequence $\{a_n\}$. This example explores this result.

Assume that the series $\sum_{k=1}^{\infty} a_k$ converges and has sum equal to L .

- What is the n th partial sum S_n of the series $\sum_{k=1}^{\infty} a_k$?
- What is the $(n - 1)$ st partial sum S_{n-1} of the series $\sum_{k=1}^{\infty} a_k$?
- What is the difference between the n th partial sum and the $(n - 1)$ st partial sum of the series $\sum_{k=1}^{\infty} a_k$?
- Since we are assuming that $\sum_{k=1}^{\infty} a_k = L$, what does that tell us about $\lim_{n \rightarrow \infty} S_n$? Why? What does that tell us about $\lim_{n \rightarrow \infty} S_{n-1}$? Why?
- Combine the results of the previous two parts of this example to determine $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} (S_n - S_{n-1})$.

Hint.

- Recall that the n th partial sum is the sum of the first n terms.
- Consider the sum with one fewer term.
- Subtract your results in (a) and (b).
- Remember that the partial sums have to tend to the sum of the series.
- Note that $S_n \rightarrow L$ and $S_{n-1} \rightarrow L$.

Answer.

- $S_n = \sum_{k=1}^n a_k$.
- $S_{n-1} = \sum_{k=1}^{n-1} a_k$.
- $\sum_{k=1}^n a_k - \sum_{k=1}^{n-1} a_k = a_n$.
- $\lim_{n \rightarrow \infty} S_n = L$ and $\lim_{n \rightarrow \infty} S_{n-1} = L$.
- We have $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} (S_n - S_{n-1}) = 0$.

Solution.

a. The n th partial sum of the series $\sum_{k=1}^{\infty} a_k$ is

$$S_n = \sum_{k=1}^n a_k.$$

b. The $(n - 1)$ st partial sum of the series $\sum_{k=1}^{\infty} a_k$ is

$$S_{n-1} = \sum_{k=1}^{n-1} a_k.$$

c. The difference between S_n and S_{n-1} is

$$\sum_{k=1}^n a_k - \sum_{k=1}^{n-1} a_k = a_n.$$

d. Since $\sum_{k=1}^{\infty} a_k = \lim_{n \rightarrow \infty} S_n$ we must have $\lim_{n \rightarrow \infty} S_n = L$. Also,

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} S_{n-1},$$

so $\lim_{n \rightarrow \infty} S_{n-1} = L$ as well.

e. We have

$$\begin{aligned} \lim_{n \rightarrow \infty} a_n &= \lim_{n \rightarrow \infty} (S_n - S_{n-1}) \\ &= \lim_{n \rightarrow \infty} S_n - \lim_{n \rightarrow \infty} S_{n-1} \\ &= L - L \\ &= 0. \end{aligned}$$

The result of Example 8.3.6 is the following important conditional statement:

If the series $\sum_{k=1}^{\infty} a_k$ converges, then the sequence $\{a_k\}$ of k th terms converges to 0.

It is logically equivalent to say that if the sequence $\{a_k\}$ does not converge to 0, then the series $\sum_{k=1}^{\infty} a_k$ cannot converge. This statement is called the Divergence Test.

The Divergence Test

If $\lim_{k \rightarrow \infty} a_k \neq 0$, then the series $\sum a_k$ diverges.

Example 8.3.7. Determine if the Divergence Test applies to the following series. If the test does not apply, explain why. If the test does apply, what does it tell us about the series?

a. $\sum \frac{k}{k+1}$

b. $\sum (-1)^k$

c. $\sum \frac{1}{k}$

Hint.

- a. Consider $\lim_{k \rightarrow \infty} \frac{k}{k+1}$.
- b. Think about the partial sums of this series.
- c. Remember that if the terms of a series go to zero, the Divergence Test doesn't make any claims.

Answer.

- a. $\sum \frac{k}{k+1}$ diverges.
- b. $\sum (-1)^k$ diverges.
- c. The Divergence Test does not apply.

Solution.

- a. Since the sequence $\frac{k}{k+1}$ converges to 1 and not 0 as k goes to infinity, the Divergence Test applies here and shows that the series $\sum \frac{k}{k+1}$ diverges.
- b. Since the terms of this series don't even converge to a number, they can't converge to zero, and thus this series diverges.
- c. Since $\lim_{k \rightarrow \infty} \frac{1}{k} = 0$, the Divergence Test does not apply to this series.

Note well: be very careful with the Divergence Test. This test only tells us what happens to a series if the terms of the corresponding sequence do not converge to 0. If the sequence of the terms of the series does converge to 0, the Divergence Test does not apply: indeed, as we will soon see, a series whose terms go to zero may either converge or diverge.

8.3.3 The Integral Test

The Divergence Test settles the questions of divergence or convergence of series $\sum a_k$ in which $\lim_{k \rightarrow \infty} a_k \neq 0$. Determining the convergence or divergence of series $\sum a_k$ in which $\lim_{k \rightarrow \infty} a_k = 0$ turns out to be more complicated. Often, we have to investigate the sequence of partial sums or apply some other technique.

Next, we consider the **harmonic series**²

$$\sum_{k=1}^{\infty} \frac{1}{k}.$$

²This series is called harmonic because each term in the series after the first is the harmonic mean of the term before it and the term after it. The harmonic mean of two numbers a and b is $\frac{2ab}{a+b}$. See “What’s Harmonic about the Harmonic Series”, by David E. Kullman (in the *College Mathematics Journal*, Vol. 32, No. 3 (May, 2001), 201-203) for an interesting discussion of the harmonic mean.

The first 9 partial sums of this series are shown following.

$$\sum_{k=1}^1 \frac{1}{k} = 1$$

$$\sum_{k=1}^4 \frac{1}{k} = 2.083333333$$

$$\sum_{k=1}^7 \frac{1}{k} = 2.592857143$$

$$\sum_{k=1}^2 \frac{1}{k} = 1.5$$

$$\sum_{k=1}^5 \frac{1}{k} = 2.283333333$$

$$\sum_{k=1}^8 \frac{1}{k} = 2.717857143$$

$$\sum_{k=1}^3 \frac{1}{k} = 1.833333333$$

$$\sum_{k=1}^6 \frac{1}{k} = 2.450000000$$

$$\sum_{k=1}^9 \frac{1}{k} = 2.828968254$$

This information alone doesn't seem to be enough to tell us if the series $\sum_{k=1}^{\infty} \frac{1}{k}$ converges or diverges. The partial sums could eventually level off to some fixed number or continue to grow without bound. Even if we look at larger partial sums, such as $\sum_{n=1}^{1000} \frac{1}{k} \approx 7.485470861$, the result isn't particularly convincing one way or another. The Integral Test is one way to determine whether or not the harmonic series converges, and we explore this test further in the next example.

Example 8.3.8. Consider the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$. Recall that the harmonic series will converge provided that its sequence of partial sums converges. The n th partial sum S_n of the series $\sum_{k=1}^{\infty} \frac{1}{k}$ is

$$\begin{aligned} S_n &= \sum_{k=1}^n \frac{1}{k} \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \\ &= 1(1) + (1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{3}\right) + \cdots + (1)\left(\frac{1}{n}\right). \end{aligned}$$

Through this last expression for S_n , we can visualize this partial sum as a sum of areas of rectangles with heights $\frac{1}{m}$ and bases of length 1, as shown in Figure 8.3.9, which uses the 9th partial sum.



Figure 8.3.9: A picture of the 9th partial sum of the harmonic series as a sum of rectangles.

The graph of the continuous function f defined by $f(x) = \frac{1}{x}$ is overlaid on this plot.

- Explain how this picture represents a particular Riemann sum.
- What is the definite integral that corresponds to the Riemann sum you considered in (a)?
- Which is larger, the definite integral in (b), or the corresponding partial sum S_9 of the series? Why?
- If instead of considering the 9th partial sum, we consider the n th partial sum, and we let n go to infinity, we can then compare the series $\sum_{k=1}^{\infty} \frac{1}{k}$ to the improper integral $\int_1^{\infty} \frac{1}{x} dx$. Which of these quantities is larger? Why?
- Does the improper integral $\int_1^{\infty} \frac{1}{x} dx$ converge or diverge? What does that result, together with your work in (d), tell us about the series $\sum_{k=1}^{\infty} \frac{1}{k}$?

Hint.

- Small hints for each of the prompts above.

Answer.

- The n th partial sum of the series $\sum_{k=1}^{\infty} \frac{1}{k}$ is the left hand Riemann sum of $f(x)$ on the interval $[1, n]$.
- $\sum_{k=1}^n \frac{1}{k} > \int_1^n \frac{1}{x} dx$.
- $\sum_{k=1}^{\infty} \frac{1}{k} > \int_1^{\infty} \frac{1}{x} dx$.
- $\int_1^{\infty} f(x) dx = \lim_{t \rightarrow \infty} \int_1^t \frac{1}{x} dx \infty$ so the series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges.

Solution.

- Notice that the n th partial sum of the series $\sum_{k=1}^{\infty} \frac{1}{k}$ is equal to the left hand Riemann sum of $f(x)$ on the interval $[1, n]$.
- Since f is a decreasing function, it follows that

$$\sum_{k=1}^n \frac{1}{k} > \int_1^n \frac{1}{x} dx.$$

- Letting Since f is decreasing, the improper integral $\int_1^{\infty} \frac{1}{x} dx$ is smaller than the limit of the Riemann sums as n goes to infinity. So we wind up with a comparison between the series $\sum_{k=1}^{\infty} \frac{1}{k}$ and the improper integral:

$$\sum_{k=1}^{\infty} \frac{1}{k} > \int_1^{\infty} \frac{1}{x} dx.$$

d. We can evaluate the improper integral as follows:

$$\begin{aligned}\int_1^\infty f(x) dx &= \lim_{t \rightarrow \infty} \int_1^t \frac{1}{x} dx \\ &= \lim_{t \rightarrow \infty} \ln(x) \Big|_1^t \\ &= \lim_{t \rightarrow \infty} (\ln(t) - \ln(1)) \\ &= \infty.\end{aligned}$$

Since the value of the series $\sum_{k=1}^\infty \frac{1}{k}$ exceeds the value of the infinite improper integral, we must conclude that the series $\sum_{k=1}^\infty \frac{1}{k}$ diverges.

The ideas from Example 8.3.8 hold more generally. Suppose that f is a continuous decreasing function and that $a_k = f(k)$ for each value of k . Consider the corresponding series $\sum_{k=1}^\infty a_k$. The partial sum

$$S_n = \sum_{k=1}^n a_k$$

can always be viewed as a left hand Riemann sum of $f(x)$, using rectangles with bases of width 1 and heights given by the values a_k . A representative picture is shown at left in Figure 8.3.10. Since f is a decreasing function, we have that

$$S_n > \int_1^n f(x) dx.$$

Taking the limit as n goes to infinity shows that

$$\sum_{k=1}^\infty a_k > \int_1^\infty f(x) dx.$$

Therefore, if the improper integral $\int_1^\infty f(x) dx$ diverges, so does the series $\sum_{k=1}^\infty a_k$.

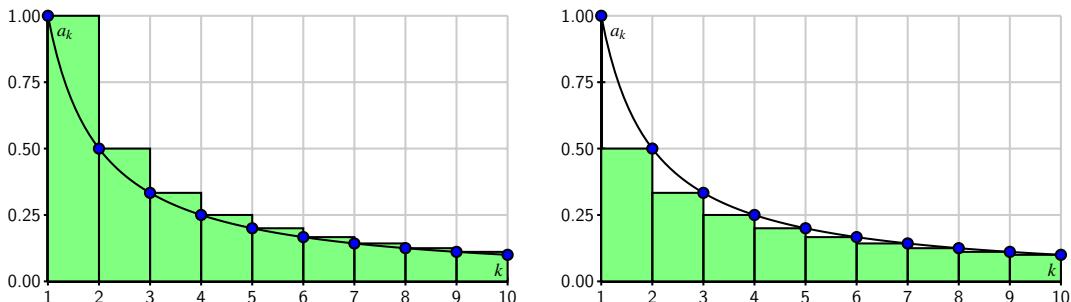


Figure 8.3.10: Comparing an improper integral to a series

What's more, if we look at the right hand Riemann sums of f on $[1, n]$ as shown at right in

Figure 8.3.10, we see that

$$\int_1^\infty f(x) dx > \sum_{k=2}^\infty a_k.$$

So if $\int_1^\infty f(x) dx$ converges, then so does $\sum_{k=2}^\infty a_k$, which also means that the series $\sum_{k=1}^\infty a_k$ also converges. Our preceding discussion has demonstrated the truth of the Integral Test.

The Integral Test

Let f be a real valued function and assume f is decreasing and positive for all x larger than some number c . Let $a_k = f(k)$ for each positive integer k .

- If the improper integral $\int_c^\infty f(x) dx$ converges, then the series $\sum_{k=1}^\infty a_k$ converges.
- If the improper integral $\int_c^\infty f(x) dx$ diverges, then the series $\sum_{k=1}^\infty a_k$ diverges.

The Integral Test compares a given infinite series to a natural, corresponding improper integral and says that the infinite series and corresponding improper integral both have the same convergence status. In the next example, we apply the Integral Test to determine the convergence or divergence of a class of important series.

Example 8.3.11. The series $\sum \frac{1}{k^p}$ are special series called p -series. We have already seen that the p -series with $p = 1$ (the harmonic series) diverges. We investigate the behavior of other p -series in this example.

- Evaluate the improper integral $\int_1^\infty \frac{1}{x^2} dx$. Does the series $\sum_{k=1}^\infty \frac{1}{k^2}$ converge or diverge? Explain.
- Evaluate the improper integral $\int_1^\infty \frac{1}{x^p} dx$ where $p > 1$. For which values of p can we conclude that the series $\sum_{k=1}^\infty \frac{1}{k^p}$ converges?
- Evaluate the improper integral $\int_1^\infty \frac{1}{x^p} dx$ where $p < 1$. What does this tell us about the corresponding p -series $\sum_{k=1}^\infty \frac{1}{k^p}$?
- Summarize your work in this example by completing the following statement.

The p -series $\sum_{k=1}^\infty \frac{1}{k^p}$ converges if and only if _____.

Hint.

- Small hints for each of the prompts above.

Answer.

- $\sum_{k=1}^\infty \frac{1}{k^2}$ converges.
- $\sum_{k=1}^\infty \frac{1}{k^p}$ converges when $p > 1$.

- c. $\sum_{k=1}^{\infty} \frac{1}{k^p}$ diverges when $p < 1$.
 d. The p -series $\sum_{k=1}^{\infty} \frac{1}{k^p}$ converges if and only if $p > 1$.

Solution.

- a. We evaluate the improper integral:

$$\begin{aligned} \int_1^{\infty} \frac{1}{x^2} dx &= \lim_{t \rightarrow \infty} \int_1^t \frac{1}{x^2} dx \\ &= \lim_{t \rightarrow \infty} -\frac{1}{x} \Big|_1^t \\ &= \lim_{t \rightarrow \infty} \left(-\frac{1}{t} + \frac{1}{1} \right) \\ &= 1. \end{aligned}$$

So the improper integral converges. The Integral Test then shows that the series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges.

- b. Assume $p > 1$. Then $p - 1 > 0$ and so $x^{-p+1} = \frac{1}{x^{p-1}}$ and

$$\lim_{x \rightarrow \infty} x^{-p+1} = \lim_{x \rightarrow \infty} \frac{1}{x^{p-1}} = 0.$$

Thus,

$$\begin{aligned} \int_1^{\infty} \frac{1}{x^p} dx &= \lim_{t \rightarrow \infty} \int_1^t \frac{1}{x^p} dx \\ &= \lim_{t \rightarrow \infty} \frac{x^{-p+1}}{-p+1} \Big|_1^t \\ &= \frac{1}{1-p} \lim_{t \rightarrow \infty} (t^{-p+1} - 1) \\ &= \frac{1}{p-1}. \end{aligned}$$

So the improper integral $\int_1^{\infty} \frac{1}{x^p} dx$ converges when $p > 1$. The Integral Test then shows that the series $\sum_{k=1}^{\infty} \frac{1}{k^p}$ converges when $p > 1$.

- c. Assume $p < 1$. Then $1 - p > 0$ and so

$$\lim_{x \rightarrow \infty} x^{-p+1} = \infty.$$

Thus,

$$\begin{aligned} \int_1^{\infty} \frac{1}{x^p} dx &= \lim_{t \rightarrow \infty} \int_1^t \frac{1}{x^p} dx \\ &= \lim_{t \rightarrow \infty} \frac{x^{-p+1}}{-p+1} \Big|_1^t \end{aligned}$$

$$= \frac{1}{1-p} \lim_{t \rightarrow \infty} (t^{-p+1} - 1) \\ = \infty.$$

So the improper integral $\int_1^\infty \frac{1}{x^p} dx$ diverges when $p < 1$. The Integral Test then shows that the series $\sum_{k=1}^\infty \frac{1}{k^p}$ also diverges when $p < 1$.

- d. We complete the statement as

The p -series $\sum_{k=1}^\infty \frac{1}{k^p}$ converges if and only if $p > 1$.

8.3.4 The Limit Comparison Test

The Integral Test allows us to determine the convergence of an entire family of series: the p -series. However, we have seen that it is often difficult to integrate functions, so the Integral Test is not one that we can use all of the time. In fact, even for a relatively simple series such as $\sum \frac{k^2+1}{k^4+2k+2}$, the Integral Test is not an option. In what follows we develop a test that applies to series of rational functions by comparing their behavior to the behavior of p -series.

Example 8.3.12. Consider the series $\sum \frac{k+1}{k^3+2}$. Since the convergence or divergence of a series only depends on the behavior of the series for large values of k , we might examine the terms of this series more closely as k gets large.

- By computing the value of $\frac{k+1}{k^3+2}$ for $k = 100$ and $k = 1000$, explain why the terms $\frac{k+1}{k^3+2}$ are essentially $\frac{k}{k^3}$ when k is large.
- Let's formalize our observations in (a) a bit more. Let $a_k = \frac{k+1}{k^3+2}$ and $b_k = \frac{k}{k^3}$. Calculate

$$\lim_{k \rightarrow \infty} \frac{a_k}{b_k}.$$

What does the value of the limit tell you about a_k and b_k for large values of k ? Compare your response from part (a).

- Does the series $\sum \frac{k}{k^3}$ converge or diverge? Why? What do you think that tells us about the convergence or divergence of the series $\sum \frac{k+1}{k^3+2}$? Explain.

Hint.

- Small hints for each of the prompts above.

Answer.

- $\frac{k+1}{k^3+2}$ looks like $\frac{k}{k^3}$ when k is large.
- $\lim_{k \rightarrow \infty} \frac{a_k}{b_k} = 1$ so $a_k \approx b_k$ for large values of k .
- $\sum \frac{k+1}{k^3+2}$ converges.

Solution.

- a. When k is very large, the constant 1 in the numerator is negligible compared to k and the constant 2 in the denominator is negligible compared to k^3 . So the numerator looks like k and the denominator k^3 when k is large. It follows that $\frac{k+1}{k^3+2}$ looks like $\frac{k}{k^3}$ when k is large.

- b. Note that

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{a_k}{b_k} &= \lim_{k \rightarrow \infty} \frac{\frac{k+1}{k^3+2}}{\frac{k}{k^3}} \\ &= \lim_{k \rightarrow \infty} \frac{(k+1)k^3}{k(k^3+2)} \\ &= \lim_{k \rightarrow \infty} \frac{1 + \frac{1}{k}}{1 + \frac{2}{k^3}} \\ &= 1.\end{aligned}$$

This tells us that $a_k \approx b_k$ for large values of k , which is what we suggested in part (a).

- c. Since $\frac{k}{k^3} = \frac{1}{k^2}$, the series $\sum \frac{k}{k^3}$ is a p -series with $p = 2$ and so converges. Since $a_k \approx b_k$ for large values of k , it seems reasonable to expect that $\sum a_k \approx \sum b_k$ for large k s. Since $\sum a_k$ is finite, we should then conclude that $\sum b_k$ is also finite. So $\sum \frac{k+1}{k^3+2}$ should be a convergent series.

Example 8.3.12 illustrates how we can compare one series with positive terms to another whose convergence status we know. Suppose we have two series $\sum a_k$ and $\sum b_k$ with positive terms and we know the convergence status of the series $\sum a_k$. Recall that the convergence or divergence of a series depends only on the terms of the series for large values of k , so if we know that a_k and b_k are proportional for large k , then the two series $\sum a_k$ and $\sum b_k$ should behave the same way. In other words, if there is a positive finite constant c such that

$$\lim_{k \rightarrow \infty} \frac{b_k}{a_k} = c,$$

then $b_k \approx ca_k$ for large values of k . So

$$\sum b_k \approx \sum ca_k = c \sum a_k.$$

Since multiplying by a nonzero constant does not affect the convergence or divergence of a series, it follows that the series $\sum a_k$ and $\sum b_k$ either both converge or both diverge. The formal statement of this fact is called the Limit Comparison Test.

The Limit Comparison Test

Let $\sum a_k$ and $\sum b_k$ be series with positive terms. If

$$\lim_{k \rightarrow \infty} \frac{b_k}{a_k} = c$$

for some positive (finite) constant c , then $\sum a_k$ and $\sum b_k$ either both converge or both diverge.

The Limit Comparison Test shows that if we have a series $\sum \frac{p(k)}{q(k)}$ of rational functions where $p(k)$ is a polynomial of degree m and $q(k)$ a polynomial of degree l , then the series $\sum \frac{p(k)}{q(k)}$ will behave like the series $\sum \frac{k^m}{k^l}$. So this test allows us to determine the convergence or divergence of series whose terms are rational functions.

Example 8.3.13. Use the Limit Comparison Test to determine the convergence or divergence of the series

$$\sum \frac{3k^2 + 1}{5k^4 + 2k + 2}.$$

by comparing it to the series $\sum \frac{1}{k^2}$.

Hint.

- a. Small hints for each of the prompts above.

Answer. $\sum \frac{k^2+1}{k^4+2k+2}$ converges.

Solution. Let $b_k = \frac{k^2+1}{k^4+2k+2}$ and $a_k = \frac{1}{k^2}$. Then

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{b_k}{a_k} &= \lim_{k \rightarrow \infty} \frac{\frac{k^2+1}{k^4+2k+2}}{\frac{1}{k^2}} \\ &= \lim_{k \rightarrow \infty} \frac{k^4 + k^2}{k^4 + 2k + 2} \\ &= \lim_{k \rightarrow \infty} \frac{1 + \frac{1}{k^2}}{1 + \frac{2}{k^3} + \frac{2}{k^4}} \\ &= 1. \end{aligned}$$

Since $\lim_{k \rightarrow \infty} \frac{b_k}{a_k}$ is a finite positive constant, the Limit Comparison Test shows that $\sum \frac{1}{k^2}$ and $\sum \frac{k^2+1}{k^4+2k+2}$ either both converge or both diverge. We know that $\sum \frac{1}{k^2}$ is a p -series with $p > 1$ and so $\sum \frac{1}{k^2}$ converges. Therefore, the series $\sum \frac{k^2+1}{k^4+2k+2}$ also converges.

8.3.5 The Ratio Test

The Limit Comparison Test works well if we can find a series with known behavior to compare. But such series are not always easy to find. In this section we will examine a test that allows us to examine the behavior of a series by comparing it to a geometric series, without knowing in advance which geometric series we need.

Example 8.3.14. Consider the series defined by

$$\sum_{k=1}^{\infty} \frac{2^k}{3^k - k}. \quad (8.3.4)$$

This series is not a geometric series, but this example will illustrate how we might compare this series to a geometric one. Recall that a series $\sum a_k$ is geometric if the ratio $\frac{a_{k+1}}{a_k}$ is always the same. For the series in (8.3.4), note that $a_k = \frac{2^k}{3^k - k}$.

- a. To see if $\sum \frac{2^k}{3^k - k}$ is comparable to a geometric series, we analyze the ratios of successive terms in the series. Complete Table 8.3.15, listing your calculations to at least 8 decimal places.

k	5	10	20	21	22	23	24	25
$\frac{a_{k+1}}{a_k}$								
a_k								

Table 8.3.15: Ratios of successive terms in the series $\sum \frac{2^k}{3^k - k}$

- b. Based on your calculations in Table 8.3.15, what can we say about the ratio $\frac{a_{k+1}}{a_k}$ if k is large?
c. Do you agree or disagree with the statement: “the series $\sum \frac{2^k}{3^k - k}$ is approximately geometric when k is large”? If not, why not? If so, do you think the series $\sum \frac{2^k}{3^k - k}$ converges or diverges? Explain.

Answer.

a.

k	5	10	20	21	22	23	24	25
$\frac{a_{k+1}}{a_k}$	0.6583679115	0.6665951585	0.6666666642	0.6666666658	0.6666666664	0.6666666666	0.66666666666	0.666666666666

b. $\frac{a_{k+1}}{a_k} \approx \frac{2}{3}$ when k is large.

c. $\sum \frac{2^k}{3^k - k}$ converges.

Solution.

- a. Ratios of successive summands in the series are shown (to 10 decimal places) in the following table.

k	5	10	20	21	22	23	24	25
$\frac{a_{k+1}}{a_k}$	0.6583679115	0.6665951585	0.6666666642	0.6666666658	0.6666666664	0.6666666666	0.66666666666	0.666666666666

- b. The calculations in the table in part (a) seem to indicate that the ratio $\frac{a_{k+1}}{a_k}$ is roughly $\frac{2}{3}$ when k is large.
- c. Since $\frac{a_{k+1}}{a_k} \approx \frac{2}{3}$ for large k , the series $\sum \frac{2^k}{3^{k-k}}$ is approximately the same as $\sum \left(\frac{2}{3}\right)^k$ when k is large. So the series $\sum \frac{2^k}{3^{k-k}}$ is approximately geometric with ratio $\frac{2}{3}$ when k is large. Since the series $\sum \left(\frac{2}{3}\right)^k$ converges because the ratio is less than 1, we expect that the series $\sum \frac{2^k}{3^{k-k}}$ will also converge.

We can generalize the argument in Example 8.3.14 in the following way. Consider the series $\sum a_k$. If

$$\frac{a_{k+1}}{a_k} \approx r$$

for large values of k , then $a_{k+1} \approx r a_k$ for large k and the series $\sum a_k$ is approximately the geometric series $\sum a r^k$ for large k . Since the geometric series with ratio r converges only for $-1 < r < 1$, we see that the series $\sum a_k$ will converge if

$$\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = r$$

for a value of r such that $|r| < 1$. This result is known as the Ratio Test.

The Ratio Test

Let $\sum a_k$ be an infinite series. Suppose

$$\lim_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} = r.$$

- If $0 \leq r < 1$, then the series $\sum a_k$ converges.
- If $1 < r$, then the series $\sum a_k$ diverges.
- If $r = 1$, then the test is inconclusive.

Note well: The Ratio Test looks at the limit of the ratio of consecutive terms of a given series; in so doing, the test is asking, “is this series approximately geometric?” If so, the test uses the limit of the ratio of consecutive terms to determine if the given series converges.

We have now encountered several tests for determining convergence or divergence of series.

- The Divergence Test can be used to show that a series diverges, but never to prove that a series converges.
- We used the Integral Test to determine the convergence status of an entire class of series, the p -series.
- The Limit Comparison Test works well for series that involve rational functions and which can therefore be compared to p -series.
- Finally, the Ratio Test allows us to compare our series to a geometric series; it is particularly useful for series that involve n th powers and factorials.

- Two other tests, the Direct Comparison Test and the Root Test, are discussed in the exercises.

One of the challenges of determining whether a series converges or diverges is finding which test answers that question.

Example 8.3.16. Determine whether each of the following series converges or diverges. Explicitly state which test you use.

a. $\sum \frac{k}{2^k}$

b. $\sum \frac{k^3+2}{k^2+1}$

c. $\sum \frac{10^k}{k!}$

d. $\sum \frac{k^3-2k^2+1}{k^6+4}$

Hint.

- a. Small hints for each of the prompts above.

Answer.

a. $\sum \frac{n}{2^n}$ converges.

b. $\sum \frac{k^3+2}{k^2+1}$ diverges.

c. $\sum \frac{10^k}{k!}$ converges.

d. $\sum \frac{k^3-2k^2+1}{k^6+4}$ converges.

Solution.

- a. Since $\lim_{n \rightarrow \infty} \frac{n}{2^n} = 0$, the Divergence Test does not apply. The Ratio Test is a good one to apply with series whose terms involve exponentials and polynomials.

In this example we have $a_{n+1} = \frac{n+1}{2^{n+1}}$ and $a_n = \frac{n}{2^n}$. So

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

As n goes to infinity, the fraction $\frac{n+1}{n}$ goes to 1 and so

$$\lim_{n \rightarrow \infty} \frac{n+1}{2n} = \frac{1}{2}.$$

So in this example, for large n we have that $\frac{a_{n+1}}{a_n} \approx \frac{1}{2}$ or that $\sum a_n$ is roughly a geometric series for large n with ratio $\frac{1}{2}$. Since geometric series with ratio $\frac{1}{2}$ converges, we can conclude that

$$\sum \frac{n}{2^n}$$

converges as well.

- b. For this series, notice that the degree of the numerator is greater than the degree of the denominator, so

$$\lim_{n \rightarrow \infty} \frac{n^3 + 2}{n^2 + 1} = \infty.$$

Since this limit is not 0, the Divergence Test shows that $\sum \frac{k^3+2}{k^2+1}$ diverges.

- c. We use the Ratio Test here with $a_k = \frac{10^k}{k!}$. Now

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} &= \lim_{k \rightarrow \infty} \frac{\frac{10^{k+1}}{(k+1)!}}{\frac{10^k}{k!}} \\ &= \lim_{k \rightarrow \infty} \frac{10^{k+1}(k!)^2}{10^k(k+1)!} \\ &= \lim_{k \rightarrow \infty} \frac{10}{k+1} \\ &= 0, \end{aligned}$$

so the Ratio Test shows that $\sum \frac{10^k}{k!}$ converges.

- d. If we ignore the highest powered terms, then the series $\sum \frac{k^3-2k^2+1}{k^6+4}$ looks like the series $\sum \frac{k^3}{k^6} = \sum \frac{1}{k^3}$. We formally apply the Limit Comparison Test as follows:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{k^3-2k^2+1}{k^6+4}}{\frac{1}{k^3}} &= \lim_{k \rightarrow \infty} \frac{k^6 - 2k^5 + k^3}{k^6 + 4} \\ &= \lim_{k \rightarrow \infty} \frac{1 - \frac{2}{k} + \frac{1}{k^3}}{1 + \frac{4}{k^6}} \\ &= 1. \end{aligned}$$

Since $\sum \frac{1}{k^3}$ is a p -series with $p = 3 > 1$, the series $\sum \frac{1}{k^3}$ converges. Consequently, the Limit Comparison Test shows that $\sum \frac{k^3-2k^2+1}{k^6+4}$ converges as well.

8.3.6 Summary

- An infinite series is a sum of the elements in an infinite sequence. In other words, an infinite series is a sum of the form

$$a_1 + a_2 + \cdots + a_n + \cdots = \sum_{k=1}^{\infty} a_k$$

where a_k is a real number for each positive integer k .

- The n th partial sum S_n of the series $\sum_{k=1}^{\infty} a_k$ is the sum of the first n terms of the series. That is,

$$S_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^n a_k.$$

- The sequence $\{S_n\}$ of partial sums of a series $\sum_{k=1}^{\infty} a_k$ tells us about the convergence or divergence of the series. In particular

- The series $\sum_{k=1}^{\infty} a_k$ converges if the sequence $\{S_n\}$ of partial sums converges. In this case we say that the series is the limit of the sequence of partial sums and write

$$\sum_{k=1}^{\infty} a_k = \lim_{n \rightarrow \infty} S_n.$$

- The series $\sum_{k=1}^{\infty} a_k$ diverges if the sequence $\{S_n\}$ of partial sums diverges.

8.3.7 Exercises

1. (Convergence of a sequence and its series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Two partial sums) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Convergence of a series and its sequence) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Convergence of an integral and a related series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. In this exercise we investigate the sequence $\{\frac{b^n}{n!}\}$ for any constant b .

- a. Use the Ratio Test to determine if the series $\sum \frac{10^k}{k!}$ converges or diverges.
- b. Now apply the Ratio Test to determine if the series $\sum \frac{b^k}{k!}$ converges for any constant b .
- c. Use your result from (b) to decide whether the sequence $\{\frac{b^n}{n!}\}$ converges or diverges. If the sequence $\{\frac{b^n}{n!}\}$ converges, to what does it converge? Explain your reasoning.

Answer.

- a. $\sum \frac{10^k}{k!}$ converges.
- b. $\sum \frac{10^k}{k!}$ converges.
- c. The sequence $\{\frac{b^n}{n!}\}$ has to converge to 0.

Solution.

- a. We use the Ratio Test with $a_n = \frac{10^n}{n!}$. First we consider the ratios of successive terms:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} &= \lim_{n \rightarrow \infty} \frac{\frac{10^{n+1}}{(n+1)!}}{\frac{10^n}{(n)!}} \\ &= \lim_{n \rightarrow \infty} \frac{10}{n+1} \\ &= 0. \end{aligned}$$

Since the limit of the ratio of successive terms is less than 1, we conclude that the series $\sum \frac{10^k}{k!}$ converges.

b. We use the Ratio Test with $a_n = \frac{10^n}{n!}$. First we consider the ratios of successive terms:

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} &= \lim_{n \rightarrow \infty} \frac{\frac{10^{n+1}}{(n+1)!}}{\frac{10^n}{n!}} \\ &= \lim_{n \rightarrow \infty} \frac{10}{n+1} \\ &= 0.\end{aligned}$$

Since the limit of the ratio of successive terms is less than 1, we conclude that the series $\sum \frac{10^k}{k!}$ converges.

c. Since the series $\sum \frac{b^n}{n!}$ converges, the Divergence Test tells us that the sequence $\left\{ \frac{b^n}{n!} \right\}$ has to converge to 0.

6. There is a test for convergence similar to the Ratio Test called the *Root Test*. Suppose we have a series $\sum a_k$ of positive terms so that $a_n \rightarrow 0$ as $n \rightarrow \infty$.

a. Assume

$$\sqrt[n]{a_n} \rightarrow r$$

as n goes to infinity. Explain why this tells us that $a_n \approx r^n$ for large values of n .

- b. Using the result of part (a), explain why $\sum a_k$ looks like a geometric series when n is big. What is the ratio of the geometric series to which $\sum a_k$ is comparable?
- c. Use what we know about geometric series to determine that values of r so that $\sum a_k$ converges if $\sqrt[n]{a_n} \rightarrow r$ as $n \rightarrow \infty$.

Answer.

- a. $\sqrt[n]{a_n} \approx r$ for large n .
 b. $\frac{a_{n+1}}{a_n} \approx r$.
 c. $0 < r < 1$.

Solution.

- a. Since $\sqrt[n]{a_n} \approx r$ for large n , it follows $a_n \approx r^n$ for large n .
 b. Because $a_n \approx r^n$ for large n , this means that the series $\sum a_n$ acts like the series $\sum r^n$ for sufficiently large n . The latter is a geometric series with ratio r between successive terms. Said differently, $\frac{a_{n+1}}{a_n} \approx \frac{r^{n+1}}{r^n} = r$.
 c. Since the terms in our series are all positive, we must have $r > 0$. Now the geometric series with ratio r converges only for $r < 1$, so the series

$$\sum a_k$$

will converge if

$$\lim_{n \rightarrow \infty} \sqrt[n]{a_n} = r$$

with $0 < r < 1$.

7. The associative and distributive laws of addition allow us to add finite sums in any order we want. That is, if $\sum_{k=0}^n a_k$ and $\sum_{k=0}^n b_k$ are finite sums of real numbers, then

$$\sum_{k=0}^n a_k + \sum_{k=0}^n b_k = \sum_{k=0}^n (a_k + b_k).$$

However, we do need to be careful extending rules like this to infinite series.

- a. Let $a_n = 1 + \frac{1}{2^n}$ and $b_n = -1$ for each nonnegative integer n .
- Explain why the series $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ both diverge.
 - Explain why the series $\sum_{k=0}^{\infty} (a_k + b_k)$ converges.
 - Explain why

$$\sum_{k=0}^{\infty} a_k + \sum_{k=0}^{\infty} b_k \neq \sum_{k=0}^{\infty} (a_k + b_k).$$

This shows that it is possible to have two divergent series $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ but yet have the series $\sum_{k=0}^{\infty} (a_k + b_k)$ converge.

- b. While part (a) shows that we cannot add series term by term in general, we can under reasonable conditions. The problem in part (a) is that we tried to add divergent series. In this exercise we will show that if $\sum a_k$ and $\sum b_k$ are convergent series, then $\sum (a_k + b_k)$ is a convergent series and

$$\sum (a_k + b_k) = \sum a_k + \sum b_k.$$

- i. Let A_n and B_n be the n th partial sums of the series $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$, respectively. Explain why

$$A_n + B_n = \sum_{k=1}^n (a_k + b_k).$$

- ii. Use the previous result and properties of limits to show that

$$\sum_{k=1}^{\infty} (a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k.$$

(Note that the starting point of the sum is irrelevant in this problem, so it doesn't matter where we begin the sum.)

- c. Use the prior result to find the sum of the series $\sum_{k=0}^{\infty} \frac{2^k + 3^k}{5^k}$.

Answer.

- a. i. $a_n = 1 + \frac{1}{2^n} \rightarrow 1 \neq 0$ and $b_n = -1 \rightarrow -1 \neq 0$.
ii. The series is geometric with $r = \frac{1}{2}$.
iii. Since the two individual series diverge, neither sum is a finite number, so it doesn't make any sense to add them.
- b. i. Note that $A_n + B_n = (a_1 + b_1) + (a_2 + b_2) + \cdots + (a_n + b_n)$.
ii. Note that $\lim_{n \rightarrow \infty} \sum_{k=1}^n (a_k + b_k) = \lim_{n \rightarrow \infty} (\sum_{k=1}^n a_k + \sum_{k=1}^n b_k)$.
- c. $\sum_{k=0}^{\infty} \frac{2^k + 3^k}{5^k} = \frac{25}{6}$.

Solution.

- a. i. Since $a_n = 1 + \frac{1}{2^n} \rightarrow 1 \neq 0$ and $b_n = -1 \rightarrow -1 \neq 0$ as $n \rightarrow \infty$, by the Divergence Test both $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ diverge.
 ii. However, the series

$$\sum_{k=0}^{\infty} (a_k + b_k) = \sum_{k=0}^{\infty} \left(1 + \frac{1}{2^k} - 1\right) = \sum_{k=0}^{\infty} \frac{1}{2^k}$$

converges because the series is geometric with $r = \frac{1}{2}$.

- iii. The issue is that it doesn't make any sense to talk about

$$\sum_{k=0}^{\infty} a_k + \sum_{k=0}^{\infty} b_k.$$

when $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ diverge. Since those series diverge, neither sum is a finite number, so it doesn't make any sense to add them.

- b. i. Let A_n and B_n be the n th partial sums of the series $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$, respectively. By definition,

$$A_n + B_n = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k.$$

Expanding these finite sums, we have

$$A_n + B_n = (a_1 + a_2 + \cdots + a_n) + (b_1 + b_2 + \cdots + b_n).$$

By the associative and commutative properties of finite sums, we can rearrange these sums in any order we wish, and also group the terms in any pairings we wish. Thus,

$$A_n + B_n = (a_1 + b_1) + (a_2 + b_2) + \cdots + (a_n + b_n),$$

and therefore

$$A_n + B_n = \sum_{k=1}^n (a_k + b_k).$$

- ii. The last equality we demonstrated above tells us that

$$\sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k.$$

In this equality, we can take the limit of both sides as $n \rightarrow \infty$. On the right side, we know that both $\lim_{n \rightarrow \infty} \sum_{k=1}^n a_k$ and $\lim_{n \rightarrow \infty} \sum_{k=1}^n b_k$ exist because the two series converge, and therefore the limit of the sum is the sum of the limits. Hence,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=1}^n (a_k + b_k) &= \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n a_k + \sum_{k=1}^n b_k \right) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k + \lim_{n \rightarrow \infty} \sum_{k=1}^n b_k. \end{aligned}$$

Thus we have shown that

$$\sum_{k=1}^{\infty} (a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k.$$

- c. Since both $\sum_{k=0}^{\infty} \left(\frac{2}{5}\right)^k$ and $\sum_{k=0}^{\infty} \left(\frac{3}{5}\right)^k$ are convergent geometric series, we can apply the result from the last exercise above to say that

$$\sum_{k=0}^{\infty} \frac{2^k + 3^k}{5^k} = \sum_{k=0}^{\infty} \left(\frac{2}{5}\right)^k + \sum_{k=0}^{\infty} \left(\frac{3}{5}\right)^k.$$

Applying the known sum of geometric series to each sum on the right, we then have

$$\sum_{k=0}^{\infty} \frac{2^k + 3^k}{5^k} = \frac{1}{1 - \frac{2}{5}} + \frac{1}{1 - \frac{3}{5}} = \frac{5}{3} + \frac{5}{2} = \frac{25}{6}.$$

8. In the Limit Comparison Test we compared the behavior of a series to one whose behavior we know. In that test we use the limit of the ratio of corresponding terms of the series to determine if the comparison is valid. In this exercise we see how we can compare two series directly, term by term, without using a limit of sequence. First we consider an example.

- a. Consider the series

$$\sum \frac{1}{k^2} \text{ and } \sum \frac{1}{k^2 + k}.$$

We know that the series $\sum \frac{1}{k^2}$ is a p -series with $p = 2 > 1$ and so $\sum \frac{1}{k^2}$ converges. In this part of the exercise we will see how to use information about $\sum \frac{1}{k^2}$ to determine information about $\sum \frac{1}{k^2 + k}$. Let $a_k = \frac{1}{k^2}$ and $b_k = \frac{1}{k^2 + k}$.

- i) Let S_n be the n th partial sum of $\sum \frac{1}{k^2}$ and T_n the n th partial sum of $\sum \frac{1}{k^2 + k}$. Which is larger, S_1 or T_1 ? Why?
ii) Recall that

$$S_2 = S_1 + a_2 \text{ and } T_2 = T_1 + b_2.$$

Which is larger, a_2 or b_2 ? Based on that answer, which is larger, S_2 or T_2 ?

- iii) Recall that

$$S_3 = S_2 + a_3 \text{ and } T_3 = T_2 + b_3.$$

Which is larger, a_3 or b_3 ? Based on that answer, which is larger, S_3 or T_3 ?

- iv) Which is larger, a_n or b_n ? Explain. Based on that answer, which is larger, S_n or T_n ?
v) Based on your response to the previous part of this exercise, what relationship do you expect there to be between $\sum \frac{1}{k^2}$ and $\sum \frac{1}{k^2 + k}$? Do you expect $\sum \frac{1}{k^2 + k}$ to converge or diverge? Why?
b. The example in the previous part of this exercise illustrates a more general result. Explain why the Direct Comparison Test, stated here, works.

The Direct Comparison Test

If

$$0 \leq b_k \leq a_k$$

for every k , then we must have

$$0 \leq \sum b_k \leq \sum a_k$$

- a. If $\sum a_k$ converges, then $\sum b_k$ converges.
- b. If $\sum b_k$ diverges, then $\sum a_k$ diverges.

Note 8.3.17. This comparison test applies only to series with nonnegative terms.

- i) Use the Direct Comparison Test to determine the convergence or divergence of the series $\sum \frac{1}{k-1}$. Hint: Compare to the harmonic series.
- ii) Use the Direct Comparison Test to determine the convergence or divergence of the series $\sum \frac{k}{k^3+1}$.

Answer.

- a. i) $S_1 = 1$ and $T_1 = \frac{1}{2}$.
 ii) $S_2 > T_2$.
 iii) $S_3 > T_3$.
 iv) $S_n > T_n$.
 v) $\sum \frac{1}{k^2} > \sum \frac{1}{k^2+k}$; $\sum \frac{1}{k^2+k}$ converges.
- b. If $\sum b_k$ diverges, then $\sum b_k$ is infinite, and anything larger must also be infinite; if $\sum a_k$ is convergent then anything smaller and positive must also be finite.
 - i) Note that $0 < \frac{1}{k} < \frac{1}{k-1}$.
 - ii) Note that $\frac{1}{k^3} > \frac{1}{k^3+1}$.

Solution.

- a. i) Since $S_1 = 1$ and $T_1 = \frac{1}{2}$, we see that $S_1 > T_1$.
 ii) Since $a_2 = \frac{1}{4}$ and $b_2 = \frac{1}{6}$, it is the case that $a_2 > b_2$. We already know that $S_1 > T_1$, so adding a larger number to S_1 than we add to T_1 makes $S_2 > T_2$.
 iii) Since $a_3 = \frac{1}{9}$ and $b_3 = \frac{1}{12}$, it is the case that $a_3 > b_3$. We already know that $S_2 > T_2$, so adding a larger number to S_2 than we add to T_2 makes $S_3 > T_3$.
 iv) Since $n^2 + n > n^2$ when n is positive, we have $\frac{1}{n^2} > \frac{1}{n^2+n}$ or $a_n > b_n$. So at each step, we are adding larger numbers to the partial sums of $\sum \frac{1}{k^2}$ than we are to the partial sums of $\sum \frac{1}{k^2+k}$, we conclude that $S_n > T_n$.
 v) Since $S_n > T_n$, we should expect $\lim_{n \rightarrow \infty} S_n > \lim_{n \rightarrow \infty} T_n$ or

$$\sum \frac{1}{k^2} > \sum \frac{1}{k^2+k}.$$

Since $\sum \frac{1}{k^2}$ is a p -series with $p = 2 > 1$, we know that $\sum \frac{1}{k^2}$ converges (and is finite). Now $\sum \frac{1}{k^2+k}$ is a sum of positive terms and is therefore positive, and is also less than a finite number ($\sum \frac{1}{k^2}$), so we expect that $\sum \frac{1}{k^2+k}$ converges as well.

- b. If we have two series $\sum a_k$ and $\sum b_k$ of positive terms with $a_k \geq b_k$ for every k , then

$$\sum a_k \geq \sum b_k > 0.$$

Now if $\sum b_k$ diverges, then $\sum b_k$ is infinite. Anything larger must also be infinite and so $\sum a_k$ must also diverge. On the other hand, if $\sum a_k$ is convergent (that is, finite), then anything smaller and positive must also be finite and so $\sum b_k$ must also converge.

- i) For any k we have $k > k - 1$, so if $k > 1$ then it follows that $0 < \frac{1}{k} < \frac{1}{k-1}$. Since the harmonic series $\sum \frac{1}{k}$ diverges, the Direct Comparison Test shows that the larger series $\sum \frac{1}{k-1}$ must also diverge.
- ii) In this case we know that $k^3 < k^3 + 1$ and so if $k > 0$ then $\frac{1}{k^3} > \frac{1}{k^3 + 1}$. It follows that $\frac{k}{k^3} > \frac{k}{k^3 + 1}$ and so the series $\sum \frac{k}{k^3 + 1}$ is bounded below by 0 and above by the series $\sum \frac{k}{k^3}$. Now

$$\frac{k}{k^3} = \frac{1}{k^2}$$

and so the series $\sum \frac{k}{k^3}$ is a p -series with $p = 2$ and converges. The Direct Comparison Test shows that the smaller series $\sum \frac{k}{k^3 + 1}$ must also converge.

8.4 Alternating Series

Motivating Questions

- What is an alternating series?
 - Under what conditions does an alternating series converge? Why?
 - How well does the n th partial sum of a convergent alternating series approximate the actual sum of the series? Why?
-

So far, we've considered series with exclusively nonnegative terms. Next, we consider series that have some negative terms. For instance, the geometric series

$$2 - \frac{4}{3} + \frac{8}{9} - \cdots + 2 \left(-\frac{2}{3}\right)^n + \cdots,$$

has $a = 2$ and $r = -\frac{2}{3}$, so that every other term alternates in sign. This series converges to

$$S = \frac{a}{1-r} = \frac{2}{1 - \left(-\frac{2}{3}\right)} = \frac{6}{5}.$$

In Example 8.4.1 and our following discussion, we investigate the behavior of similar series where consecutive terms have opposite signs.

Example 8.4.1. Example 8.3.1 showed how we can approximate the number e with linear, quadratic, and other polynomial approximations. We use a similar approach in this example to obtain linear and quadratic approximations to $\ln(2)$. Along the way, we encounter a type of series that is different than most of the ones we have seen so far. Throughout this example, let $f(x) = \ln(1 + x)$.

- Find the tangent line to f at $x = 0$ and use this linearization to approximate $\ln(2)$. That is, find $L(x)$, the tangent line approximation to $f(x)$, and use the fact that $L(1) \approx f(1)$ to estimate $\ln(2)$.
- The linearization of $\ln(1 + x)$ does not provide a very good approximation to $\ln(2)$ since 1 is not that close to 0. To obtain a better approximation, we alter our approach; instead of using a straight line to approximate $\ln(2)$, we use a quadratic function to account for the concavity of $\ln(1 + x)$ for x close to 0. With the linearization, both the function's value and slope agree with the linearization's value and slope at $x = 0$. We will now make a quadratic approximation $P_2(x)$ to $f(x) = \ln(1 + x)$ centered at $x = 0$ with the property that $P_2(0) = f(0)$, $P_2'(0) = f'(0)$, and $P_2''(0) = f''(0)$.
 - Let $P_2(x) = x - \frac{x^2}{2}$. Show that $P_2(0) = f(0)$, $P_2'(0) = f'(0)$, and $P_2''(0) = f''(0)$. Use $P_2(x)$ to approximate $\ln(2)$ by using the fact that $P_2(1) \approx f(1)$.

- ii. We can continue approximating $\ln(2)$ with polynomials of larger degree whose derivatives agree with those of f at 0. This makes the polynomials fit the graph of f better for more values of x around 0. For example, let $P_3(x) = x - \frac{x^2}{2} + \frac{x^3}{3}$. Show that $P_3(0) = f(0)$, $P'_3(0) = f'(0)$, $P''_3(0) = f''(0)$, and $P'''_3(0) = f'''(0)$. Taking a similar approach to preceding questions, use $P_3(x)$ to approximate $\ln(2)$.
- iii. If we used a degree 4 or degree 5 polynomial to approximate $\ln(1+x)$, what approximations of $\ln(2)$ do you think would result? Use the preceding questions to conjecture a pattern that holds, and state the degree 4 and degree 5 approximation.

Solution.

- a. The linearization of f at $x = a$ is

$$f(a) + f'(a)(x - a),$$

so the linearization $P_1(x)$ of $f(x) = \ln(1+x)$ at $x = 0$ is

$$P_1(x) = 0 + \frac{1}{1+0}(x - 0) = x.$$

Now

$$f(x) \approx P_1(x)$$

for x close to 0 and so

$$\ln(2) = \ln(1+1) \approx P_1(1) = 1.$$

- b. i. The derivatives of P_2 and f are

$$\begin{array}{ll} P_2(x) = x - \frac{x^2}{2} & f(x) = \ln(1+x) \\ P'_2(x) = 1 - x & f'(x) = \frac{1}{1+x} \\ P''_2(x) = -1 & f''(x) = -\frac{1}{(1+x)^2}, \end{array}$$

and so the derivatives of P_2 and f evaluated at 0 are

$$\begin{array}{ll} P_2(0) = 0 & f(0) = \ln(1) = 0 \\ P'_2(0) = 1 & f'(0) = \frac{1}{1+0} = 1 \\ P''_2(0) = -1 & = -\frac{1}{(1+0)^2} = -1. \end{array}$$

Then

$$\ln(2) = \ln(1+1) \approx P_2(1) = 1 - \frac{1}{2} = \frac{1}{2}.$$

ii. The derivatives of P_3 and f are

$$\begin{array}{ll} P_3(x) = x - \frac{x^2}{2} + \frac{x^3}{3} & f(x) = \ln(1+x) \\ P'_3(x) = 1 - x + x^2 & f'(x) = \frac{1}{1+x} \\ P''_3(x) = -1 + 2x & f''(x) = -\frac{1}{(1+x)^2} \\ P'''_3(x) = 2 & f'''(x) = \frac{2}{(1+x)^3}, \end{array}$$

and so the derivatives of P_3 and f evaluated at 0 are

$$\begin{array}{ll} P_3(0) = 0 & f(0) = \ln(1+0) = 0 \\ P'_3(0) = 1 & f'(0) = \frac{1}{1+0} = 1 \\ P''_3(0) = -1 & f''(0) = -\frac{1}{(1+0)^2} = -1 \\ P'''_3(0) = 2 & f'''(0) = \frac{2}{(1+0)^3} = 2. \end{array}$$

Then

$$\ln(2) = \ln(1+1) \approx P_3(0) = 1 - \frac{1}{2} + \frac{1}{3} = \frac{5}{6} \approx 0.83.$$

8.4.1 The Alternating Series Test

Example 8.4.1 gives us several approximations to $\ln(2)$. The linear approximation is 1, and the quadratic approximation is $1 - \frac{1}{2} = \frac{1}{2}$. If we continue this process, cubic, quartic (degree 4), quintic (degree 5), and higher degree polynomials give us the approximations to $\ln(2)$ in Table 8.4.2.

linear	1	1
quadratic	$1 - \frac{1}{2}$	0.5
cubic	$1 - \frac{1}{2} + \frac{1}{3}$	$0.8\bar{3}$
quartic	$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4}$	$0.58\bar{3}$
quintic	$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5}$	$0.78\bar{3}$

Table 8.4.2

The pattern here shows that $\ln(2)$ can be approximated by the partial sums of the infinite series

$$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k} \tag{8.4.1}$$

where the alternating signs are indicated by the factor $(-1)^{k+1}$. We call such a series an *alternating series*.

Using computational technology, we find that the sum of the first 100 terms in this series

is 0.6881721793. As a comparison, $\ln(2) \approx 0.6931471806$. This shows that even though the series (8.4.1) converges to $\ln(2)$, it must do so quite slowly, since the sum of the first 100 terms isn't particularly close to $\ln(2)$. We will investigate the issue of how quickly an alternating series converges later in this section.

Definition 8.4.3. An alternating series is a series of the form

$$\sum_{k=0}^{\infty} (-1)^k a_k,$$

where $a_k > 0$ for each k .

We have some flexibility in how we write an alternating series; for example, the series

$$\sum_{k=1}^{\infty} (-1)^{k+1} a_k,$$

whose index starts at $k = 1$, is also alternating. As we will soon see, there are several very nice results that hold for alternating series, while alternating series can also demonstrate some unusual behavior.

It is important to remember that most of the series tests we have seen in previous sections apply only to series with nonnegative terms. Alternating series require a different test.

Example 8.4.4. Remember that, by definition, a series converges if and only if its corresponding sequence of partial sums converges.

- a. Calculate the first few partial sums (to 10 decimal places) of the alternating series

$$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k}.$$

Label each partial sum with the notation $S_n = \sum_{k=1}^n (-1)^{k+1} \frac{1}{k}$ for an appropriate choice of n .

- b. Plot the sequence of partial sums from part (a). What do you notice about this sequence?

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a.

$\sum_{k=1}^1 (-1)^{k+1} \frac{1}{k}$	1	$\sum_{k=1}^6 (-1)^{k+1} \frac{1}{k}$	0.6166666667
$\sum_{k=1}^2 (-1)^{k+1} \frac{1}{k}$	0.5	$\sum_{k=1}^7 (-1)^{k+1} \frac{1}{k}$	0.7595238095
$\sum_{k=1}^3 (-1)^{k+1} \frac{1}{k}$	0.8333333333	$\sum_{k=1}^8 (-1)^{k+1} \frac{1}{k}$	0.6345238095
$\sum_{k=1}^4 (-1)^{k+1} \frac{1}{k}$	0.5833333333	$\sum_{k=1}^9 (-1)^{k+1} \frac{1}{k}$	0.7456349206
$\sum_{k=1}^5 (-1)^{k+1} \frac{1}{k}$	0.7833333333	$\sum_{k=1}^{10} (-1)^{k+1} \frac{1}{k}$	0.6456349206

- b. There appears to be a limit for the sequence of partial sums.

Solution.

- a. The completed table is shown below.

$\sum_{k=1}^1 (-1)^{k+1} \frac{1}{k}$	1	$\sum_{k=1}^6 (-1)^{k+1} \frac{1}{k}$	0.6166666667
$\sum_{k=1}^2 (-1)^{k+1} \frac{1}{k}$	0.5	$\sum_{k=1}^7 (-1)^{k+1} \frac{1}{k}$	0.7595238095
$\sum_{k=1}^3 (-1)^{k+1} \frac{1}{k}$	0.8333333333	$\sum_{k=1}^8 (-1)^{k+1} \frac{1}{k}$	0.6345238095
$\sum_{k=1}^4 (-1)^{k+1} \frac{1}{k}$	0.5833333333	$\sum_{k=1}^9 (-1)^{k+1} \frac{1}{k}$	0.7456349206
$\sum_{k=1}^5 (-1)^{k+1} \frac{1}{k}$	0.7833333333	$\sum_{k=1}^{10} (-1)^{k+1} \frac{1}{k}$	0.6456349206

- b. The entries in the sequence of partial sums from part (a) is shown in the figure below.

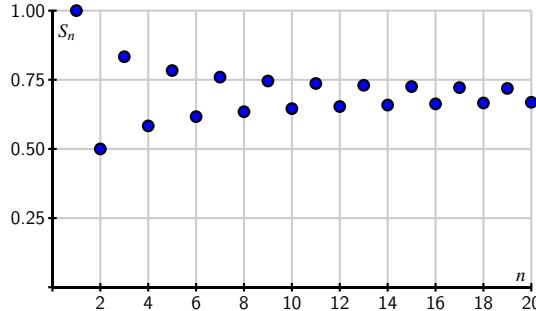


Figure 8.4.5: The sequence of partial sums.

Notice that the partial sums seem to oscillate back and forth around some fixed number, getting closer to that fixed number at each successive step. So there appears to be a limit for this sequence of partial sums.

Example 8.4.4 illustrates the general behavior of any convergent alternating series. We see that the partial sums of the alternating harmonic series oscillate around a fixed number that turns out to be the sum of the series.

Recall that if $\lim_{k \rightarrow \infty} a_k \neq 0$, then the series $\sum a_k$ diverges by the Divergence Test. From this point forward, we will thus only consider alternating series

$$\sum_{k=1}^{\infty} (-1)^{k+1} a_k$$

in which the sequence a_k consists of positive numbers that decrease to 0. The n th partial sum S_n is

$$S_n = \sum_{k=1}^n (-1)^{k+1} a_k.$$

Notice that

- $S_2 = a_1 - a_2$, and since $a_1 > a_2$ we have $0 < S_2 < S_1$.
- $S_3 = S_2 + a_3$ and so $S_2 < S_3$. But $a_3 < a_2$, so $S_3 < S_1$. Thus, $0 < S_2 < S_3 < S_1$.
- $S_4 = S_3 - a_4$ and so $S_4 < S_3$. But $a_4 < a_3$, so $S_2 < S_4$. Thus, $0 < S_2 < S_4 < S_3 < S_1$.
- $S_5 = S_4 + a_5$ and so $S_4 < S_5$. But $a_5 < a_4$, so $S_5 < S_3$. Thus, $0 < S_2 < S_4 < S_5 < S_3 < S_1$.

This pattern continues as illustrated in Figure 8.4.6 (with n odd) so that each partial sum lies between the previous two partial sums.

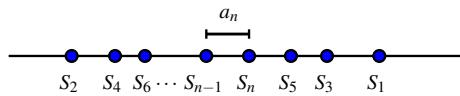


Figure 8.4.6: Partial sums of an alternating series

Note further that the absolute value of the difference between the $(n - 1)$ st partial sum S_{n-1} and the n th partial sum S_n is

$$|S_n - S_{n-1}| = a_n.$$

Because the sequence $\{a_n\}$ converges to 0, the distance between successive partial sums becomes as close to zero as we'd like, and thus the sequence of partial sums converges (even though we don't know the exact value to which it converges).

The preceding discussion has demonstrated the truth of the Alternating Series Test.

The Alternating Series Test

Given an alternating series $\sum (-1)^k a_k$, if the sequence $\{a_k\}$ of positive terms decreases to 0 as $k \rightarrow \infty$, then the alternating series converges.

Note that if the limit of the sequence $\{a_k\}$ is not 0, then the alternating series diverges.

Example 8.4.7. Which series converge and which diverge? Justify your answers.

a. $\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2 + 2}$

b. $\sum_{k=1}^{\infty} \frac{(-1)^{k+1} 2k}{k + 5}$

c. $\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$

Hint.

- a. Small hints for each of the prompts above.

Answer.

a. $\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2+2}$ converges.

b. $\sum_{k=1}^{\infty} \frac{2(-1)^{k+1}k}{k+5}$ diverges.

c. $\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$ converges.

Solution.

- a. Since $(k+1)^2 + 2 > k^2 + 2$ it follows that $\frac{1}{(k+1)^2+2} < \frac{1}{k^2+2}$ and the sequence $\frac{1}{k^2+2}$ decreases to 0. The Alternating Series Test shows then that the series $\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2+2}$ converges.
- b. In this case we have that $\lim_{k \rightarrow \infty} \frac{2k}{k+5} = 2$ and so the sequence of k th terms does not even converge to 0. So the series $\sum_{k=1}^{\infty} \frac{2(-1)^{k+1}k}{k+5}$ diverges by the Divergence Test.
- c. We know that the natural log is an increasing function, so $\frac{1}{\ln(k)}$ is a decreasing function. It is also true that $\lim_{k \rightarrow \infty} \frac{1}{\ln(k)} = 0$ and so the series $\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$ converges by the Alternating Series Test.

8.4.2 Estimating Alternating Sums

If the series converges, the argument for the Alternating Series Test also provides us with a method to determine how close the n th partial sum S_n is to the actual sum of the series. To see how this works, let S be the sum of a convergent alternating series, so

$$S = \sum_{k=1}^{\infty} (-1)^k a_k.$$

Recall that the sequence of partial sums oscillates around the sum S so that

$$|S - S_n| < |S_{n+1} - S_n| = a_{n+1}.$$

Therefore, the value of the term a_{n+1} provides an error estimate for how well the partial sum S_n approximates the actual sum S . We summarize this fact in the statement of the Alternating Series Estimation Theorem.

Alternating Series Estimation Theorem

If the alternating series $\sum_{k=1}^{\infty} (-1)^{k+1} a_k$ converges and has sum S , and $S_n = \sum_{k=1}^n (-1)^{k+1} a_k$ is the n th partial sum of the alternating series, then

$$\left| \sum_{k=1}^{\infty} (-1)^{k+1} a_k - S_n \right| \leq a_{n+1}.$$

Example 8.4.8. Determine how well the 100th partial sum S_{100} of

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$$

approximates the sum of the series.

Solution. If we let S be the sum of the series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$, then we know that

$$|S_{100} - S| < a_{101}.$$

Now

$$a_{101} = \frac{1}{101} \approx 0.0099,$$

so the 100th partial sum is within 0.0099 of the sum of the series. We have discussed the fact (and will later verify) that

$$S = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln(2),$$

and so $S \approx 0.693147$ while

$$S_{100} = \sum_{k=1}^{100} \frac{(-1)^{k+1}}{k} \approx 0.6881721793.$$

We see that the actual difference between S and S_{100} is approximately 0.0049750013, which is indeed less than 0.0099.

Example 8.4.9. Determine the number of terms it takes to approximate the sum of the convergent alternating series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^4}$$

to within 0.0001.

Hint. Small hints for each of the prompts above.

Answer. $S_{10} \approx 0.9469925924$; $\frac{7}{720} \pi^4 \approx 0.9470328299$.

Solution. First note that the sequence $\left\{\frac{1}{k^4}\right\}$ decreases to 0, so the series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^4}$ converges by the Alternating Series Test. Let S be the sum $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^4}$ and S_n the n th partial sum of the series. We know that

$$|S - S_n| < a_{n+1} = \frac{1}{(n+1)^4}.$$

So we only need to determine the value of n so that

$$\frac{1}{(n+1)^4} < 0.0001.$$

This happens when $(n+1)^4 < 10000$ or when $n+1 > 10$. So $n = 10$ will do. A computer algebra system gives $S_{10} \approx 0.9469925924$ while $\frac{7}{720}\pi^4 \approx 0.9470328299$. These do agree to 0.0001.

8.4.3 Absolute and Conditional Convergence

A series such as

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots \quad (8.4.2)$$

whose terms are neither all nonnegative nor alternating is different from any series that we have considered so far. The behavior of such a series can be rather complicated, but there is an important connection between a series with some negative terms and series with all positive terms.

Example 8.4.10.

- a. Explain why the series

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots$$

must have a sum that is less than the series

$$\sum_{k=1}^{\infty} \frac{1}{k^2}.$$

- b. Explain why the series

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots$$

must have a sum that is greater than the series

$$\sum_{k=1}^{\infty} -\frac{1}{k^2}.$$

c. Given that the terms in the series

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots$$

converge to 0, what do you think the previous two results tell us about the convergence status of this series?

Hint.

a. Small hints for each of the prompts above.

Answer.

a. $1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots < \sum_{k=1}^{\infty} \frac{1}{k^2}$.

b. $1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots > \sum_{k=1}^{\infty} -\frac{1}{k^2}$.

c. We expect this series to converge to some finite number between $\sum_{k=1}^{\infty} -\frac{1}{k^2}$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$.

Solution.

a. Each term in the series

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots$$

is of the form $\frac{1}{k^2}$ or $-\frac{1}{k^2}$. Since each of these terms is less than or equal to $\frac{1}{k^2}$ it follows that

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots < \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

b. Each term in the series

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots$$

is of the form $\frac{1}{k^2}$ or $-\frac{1}{k^2}$. Since each of these terms is greater than or equal to $-\frac{1}{k^2}$ it follows that

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots > \sum_{k=1}^{\infty} -\frac{1}{k^2}.$$

c. We know that $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is a p -series with $p = 2 > 1$ and so is a convergent series. Since $\sum_{k=1}^{\infty} \frac{1}{k^2} = -\sum_{k=1}^{\infty} -\frac{1}{k^2}$, the series $\sum_{k=1}^{\infty} -\frac{1}{k^2}$ also converges. Since the terms in the series

$$1 - \frac{1}{4} - \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} - \frac{1}{49} - \frac{1}{64} - \frac{1}{81} - \frac{1}{100} + \dots$$

converge to 0 and the series itself is bounded between two convergent series, we should expect this series to converge to some finite number between $\sum_{k=1}^{\infty} -\frac{1}{k^2}$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$.

As the example in Example 8.4.10 suggests, if a series $\sum a_k$ has some negative terms but $\sum |a_k|$ converges, then the original series, $\sum a_k$, must also converge. That is, if $\sum |a_k|$ converges, then so must $\sum a_k$.

As we just observed, this is the case for the series (8.4.2), because the corresponding series of the absolute values of its terms is the convergent p -series $\sum \frac{1}{k^2}$. But there are series, such as the alternating harmonic series $\sum (-1)^{k+1} \frac{1}{k}$, that converge while the corresponding series of absolute values, $\sum \frac{1}{k}$, diverges. We distinguish between these behaviors by introducing the following language.

Definition 8.4.11. Consider a series $\sum a_k$.

- a. The series $\sum a_k$ converges *absolutely* (or is *absolutely convergent*) provided that $\sum |a_k|$ converges.
- b. The series $\sum a_k$ converges *conditionally* (or is *conditionally convergent*) provided that $\sum |a_k|$ diverges and $\sum a_k$ converges.

In this terminology, the series (8.4.2) converges absolutely while the alternating harmonic series is conditionally convergent.

Example 8.4.12.

- a. Consider the series $\sum (-1)^k \frac{\ln(k)}{k}$.
 - i. Does this series converge? Explain.
 - ii. Does this series converge absolutely? Explain what test you use to determine your answer.
- b. Consider the series $\sum (-1)^k \frac{\ln(k)}{k^2}$.
 - i. Does this series converge? Explain.
 - ii. Does this series converge absolutely? Hint: Use the fact that $\ln(k) < \sqrt{k}$ for large values of k and then compare to an appropriate p -series.

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a. i. $\sum (-1)^k \frac{\ln(k)}{k}$ converges.
- ii. $\sum (-1)^k \frac{\ln(k)}{k}$ converges conditionally.
 - i. $\sum (-1)^k \frac{\ln(k)}{k^2}$ converges.
 - ii. $\sum (-1)^k \frac{\ln(k)}{k^2}$ converges absolutely.

Solution.

- a. i. By L'Hopital's Rule we have

$$\lim_{k \rightarrow \infty} \frac{\ln(k)}{k} = \lim_{k \rightarrow \infty} \frac{1}{k} = 0.$$

Also, $\frac{d}{dk} \frac{\ln(k)}{k} = \frac{1-\ln(k)}{k^2}$ is negative when $k > e$, so the sequence $\left\{ \frac{\ln(k)}{k} \right\}$ ultimately decreases to 0. Since the first few terms in a series are irrelevant to its convergence or divergence, we conclude that the series $\sum (-1)^k \frac{\ln(k)}{k}$ converges by the Alternating Series test.

- ii. Note that

$$\begin{aligned} \lim_{t \rightarrow \infty} \int_1^t \frac{\ln(x)}{x} dx &= \lim_{t \rightarrow \infty} \frac{\ln(x)^2}{2} \Big|_1^t \\ &= \lim_{t \rightarrow \infty} \frac{\ln(t)^2}{2} \\ &= \lim_{t \rightarrow \infty} \frac{\ln(t)^2}{2} \\ &= \infty. \end{aligned}$$

Since the improper integral diverges, the Integral Test shows that the series $\sum (-1)^k \frac{\ln(k)}{k}$ diverges. So the series $\sum (-1)^k \frac{\ln(k)}{k}$ converges conditionally.

- i. By L'Hopital's Rule we have

$$\lim_{k \rightarrow \infty} \frac{\ln(k)}{k^2} = \lim_{k \rightarrow \infty} \frac{1}{2k^2} = 0.$$

Also,

$$\frac{d}{dk} \frac{\ln(k)}{k^2} = \frac{k - 2k \ln(k)}{k^4} = \frac{1 - 2 \ln(k)}{k^3}$$

is negative when $k > e$, so the sequence $\left\{ \frac{\ln(k)}{k^2} \right\}$ ultimately decreases to 0. Since the first few terms in a series are irrelevant to its convergence or divergence, we conclude that the series $\sum (-1)^k \frac{\ln(k)}{k^2}$ converges by the Alternating Series test.

- ii. Notice that

$$\lim_{k \rightarrow \infty} \frac{\ln(k)}{k^{1/2}} = \lim_{k \rightarrow \infty} \frac{2}{k^{1/2}} = 0,$$

So $\frac{1}{\sqrt{k}}$ dominates $\ln(k)$ and $\ln(k) < \sqrt{k}$ for large k . It follows that

$$\frac{\ln(k)}{k^2} < \frac{\sqrt{k}}{k^2} = \frac{1}{k^{3/2}}$$

for large k . Therefore,

$$\sum \frac{\ln(k)}{k^2} < \sum \frac{1}{k^{3/2}}$$

for large k . Since $\sum \frac{1}{k^{3/2}}$ is a p -series with $p = \frac{3}{2} > 1$, the series $\sum \frac{1}{k^{3/2}}$ converges. This forces the series $\sum \frac{\ln(k)}{k^2}$ to converge as well. So $\sum (-1)^k \frac{\ln(k)}{k^2}$ converges absolutely.

Conditionally convergent series turn out to be very interesting. If the sequence $\{a_n\}$ decreases to 0, but the series $\sum a_k$ diverges, the conditionally convergent series $\sum (-1)^k a_k$ is right on the borderline of being a divergent series. As a result, any conditionally convergent series converges very slowly. Furthermore, some very strange things can happen with conditionally convergent series, as illustrated in some of the exercises.

8.4.4 Summary of Tests for Convergence of Series

We have discussed several tests for convergence/divergence of series in our sections and in exercises. We close this section of the text with a summary of all the tests we have encountered, followed by an example that challenges you to decide which convergence test to apply to several different series.

Geometric Series The geometric series $\sum ar^k$ with ratio r converges for $-1 < r < 1$ and diverges for $|r| \geq 1$.

The sum of the convergent geometric series $\sum_{k=0}^{\infty} ar^k$ is $\frac{a}{1-r}$.

Divergence Test If the sequence a_n does not converge to 0, then the series $\sum a_k$ diverges.

This is the first test to apply because the conclusion is simple. However, if $\lim_{n \rightarrow \infty} a_n = 0$, no conclusion can be drawn.

Integral Test Let f be a positive, decreasing function on an interval $[c, \infty)$ and let $a_k = f(k)$ for each positive integer $k \geq c$.

- If $\int_c^{\infty} f(t) dt$ converges, then $\sum a_k$ converges.
- If $\int_c^{\infty} f(t) dt$ diverges, then $\sum a_k$ diverges.

Use this test when $f(x)$ is easy to integrate.

Direct Comp. Test (see Ex 4 in Section 8.3)

Let $0 \leq a_k \leq b_k$ for each positive integer k .

- If $\sum b_k$ converges, then $\sum a_k$ converges.
- If $\sum a_k$ diverges, then $\sum b_k$ diverges.

Use this test when you have a series with known behavior that you can compare to — this test can be difficult to apply.

Limit Comp. Test Let a_n and b_n be sequences of positive terms. If

$$\lim_{k \rightarrow \infty} \frac{a_k}{b_k} = L$$

for some positive finite number L , then the two series $\sum a_k$ and $\sum b_k$ either both converge or both diverge.

Easier to apply in general than the comparison test, but you must have a series with known behavior to compare. Useful to apply to series of rational functions.

Ratio Test Let $a_k \neq 0$ for each k and suppose

$$\lim_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} = r.$$

- If $r < 1$, then the series $\sum a_k$ converges absolutely.
- If $r > 1$, then the series $\sum a_k$ diverges.
- If $r = 1$, then test is inconclusive.

This test is useful when a series involves factorials and powers.

Root Test (see Exercise 2 in Section 8.3)

Let $a_k \geq 0$ for each k and suppose

$$\lim_{k \rightarrow \infty} \sqrt[k]{a_k} = r.$$

- If $r < 1$, then the series $\sum a_k$ converges.
- If $r > 1$, then the series $\sum a_k$ diverges.
- If $r = 1$, then test is inconclusive.

In general, the Ratio Test can usually be used in place of the Root Test. However, the Root Test can be quick to use when a_k involves k th powers.

Alt. Series Test If a_n is a positive, decreasing sequence so that $\lim_{n \rightarrow \infty} a_n = 0$, then the alternating series $\sum (-1)^{k+1} a_k$ converges.

This test applies only to alternating series — we assume that the terms a_n are all positive and that the sequence $\{a_n\}$ is decreasing.

Alt. Series Est. Let $S_n = \sum_{k=1}^n (-1)^{k+1} a_k$ be the n th partial sum of the alternating series $\sum_{k=1}^{\infty} (-1)^{k+1} a_k$.

Assume $a_n > 0$ for each positive integer n , the sequence a_n decreases to 0 and $\lim_{n \rightarrow \infty} S_n = S$. Then it follows that $|S - S_n| < a_{n+1}$.

This bound can be used to determine the accuracy of the partial sum S_n as an approximation of the sum of a convergent alternating series.

Example 8.4.13. For (a)-(j), use appropriate tests to determine the convergence or divergence of the following series. Throughout, if a series is a convergent geometric series, find its sum.

a. $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k-2}}$

b. $\sum_{k=1}^{\infty} \frac{k}{1+2k}$

c. $\sum_{k=0}^{\infty} \frac{2k^2+1}{k^3+k+1}$

d.
$$\sum_{k=0}^{\infty} \frac{100^k}{k!}$$

e.
$$\sum_{k=1}^{\infty} \frac{2^k}{5^k}$$

f.
$$\sum_{k=1}^{\infty} \frac{k^3 - 1}{k^5 + 1}$$

g.
$$\sum_{k=2}^{\infty} \frac{3^{k-1}}{7^k}$$

h.
$$\sum_{k=2}^{\infty} \frac{1}{k^k}$$

i.
$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{\sqrt{k+1}}$$

j.
$$\sum_{k=2}^{\infty} \frac{1}{k \ln(k)}$$

- k. Determine a value of n so that the n th partial sum S_n of the alternating series $\sum_{n=2}^{\infty} \frac{(-1)^n}{\ln(n)}$ approximates the sum to within 0.001.

Hint.

- a. Small hints for each of the prompts above.

Answer.

a. $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k-2}}$ diverges.

b. $\sum_{k=1}^{\infty} \frac{k}{1+2k}$ diverges.

c. $\sum_{k=0}^{\infty} \frac{2k^2+1}{k^3+k+1}$ diverges.

d. $\sum_{k=0}^{\infty} \frac{100^k}{k!}$ converges.

e. $\sum_{k=1}^{\infty} \frac{2^k}{5^k}$ is a geometric series with ratio $\frac{2}{5}$ and sum $\frac{2}{3}$.

f. $\sum_{k=1}^{\infty} \frac{k^3-1}{k^5+1}$ converges.

g. $\sum_{k=2}^{\infty} \frac{3^{k-1}}{7^k}$ is a convergent geometric series with $a = \frac{3}{49}$ and $r = \frac{3}{7}$ and sum $\frac{3}{28}$.

h. $\sum_{k=2}^{\infty} \frac{1}{k^k}$ converges.

- i. $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{\sqrt{k+1}}$ converges.
- j. $\sum_{k=2}^{\infty} \frac{1}{k \ln(k)}$ diverges.
- k. $\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$ converges very slowly.

Solution.

- a. For large values of k , $\frac{2}{\sqrt{k-2}}$ looks like $\frac{2}{\sqrt{k}}$. We will use the Limit Comparison Test to compare these two series:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{2}{\sqrt{k-2}}}{\frac{2}{\sqrt{k}}} &= \lim_{k \rightarrow \infty} \frac{\sqrt{k}}{\sqrt{k-2}} \\ &= \lim_{k \rightarrow \infty} \sqrt{\frac{1}{1 - \frac{2}{k}}} \\ &= 1. \end{aligned}$$

Since this limit is a positive constant, we know that the two series $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k-2}}$ and $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k}}$ either both converge or both diverge. Now $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k}} = 2 \sum_{k=3}^{\infty} \frac{1}{k^{1/2}}$ is constant times a p -series with $p = \frac{1}{2}$, so $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k}}$ diverges. Therefore, $\sum_{k=3}^{\infty} \frac{2}{\sqrt{k-2}}$ diverges as well.

- b. Note that

$$\lim_{k \rightarrow \infty} \frac{k}{1+2k} = \lim_{k \rightarrow \infty} \frac{k}{2k} = \frac{1}{2},$$

so the Divergence Test shows that $\sum_{k=1}^{\infty} \frac{k}{1+2k}$ diverges.

- c. For large values of k , $\frac{2k^2+1}{k^3+k+1}$ looks like $\frac{2}{k}$. We will use the Limit Comparison Test to compare these two series:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{2k^2+1}{k^3+k+1}}{\frac{2}{k}} &= \lim_{k \rightarrow \infty} \frac{(2k^2+1)k}{2(k^3+k+1)} \\ &= \lim_{k \rightarrow \infty} \frac{2 + \frac{1}{k^2}}{2 + \frac{1}{k^2} + \frac{1}{k^3}} \\ &= 1. \end{aligned}$$

Since this limit is a positive constant, we know that the two series $\sum_{k=1}^{\infty} \frac{2k^2+1}{k^3+k+1}$ and $\sum_{k=1}^{\infty} \frac{2}{k}$ either both converge or both diverge. Now $\sum_{k=1}^{\infty} \frac{2}{k} = 2 \sum_{k=1}^{\infty} \frac{1}{k}$ is constant times a p -series with $p = 1$, so $\sum_{k=1}^{\infty} \frac{2}{k}$ diverges. Therefore, $\sum_{k=1}^{\infty} \frac{2k^2+1}{k^3+k+1}$ diverges as well.

- d. Series that involve factorials are good candidates for the Ratio Test:

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{\frac{100^{k+1}}{(k+1)!}}{\frac{100^k}{k!}} &= \lim_{k \rightarrow \infty} \frac{100^{k+1}k!}{100^k(k+1)!} \\ &= \lim_{k \rightarrow \infty} \frac{100}{k+1} \\ &= 0.\end{aligned}$$

Since this limit is 0, the Ratio Test tells us that the series $\sum_{k=0}^{\infty} \frac{100^k}{k!}$ converges.

- e. Notice that $\sum_{k=1}^{\infty} \frac{2^k}{5^k} = \sum_{k=1}^{\infty} \left(\frac{2}{5}\right)^k$ is a geometric series with ratio $\frac{2}{5}$. Let's write out the first few terms to identify the value of a :

$$\sum_{k=1}^{\infty} \left(\frac{2}{5}\right)^k = \frac{2}{5} + \left(\frac{2}{5}\right)^2 + \left(\frac{2}{5}\right)^3 + \cdots = \left(\frac{2}{5}\right) \left[1 + \frac{2}{5} + \left(\frac{2}{5}\right)^2 + \cdots\right].$$

So $a = \frac{2}{5}$. Since the ratio of this geometric series is between -1 and 1 , the series converges. The sum of the series is

$$\sum_{k=1}^{\infty} \frac{2^k}{5^k} = \left(\frac{2}{5}\right) \left(\frac{1}{1 - \frac{2}{5}}\right) = \left(\frac{2}{5}\right) \left(\frac{5}{3}\right) = \frac{2}{3}.$$

- f. For large values of n , $\frac{k^3-1}{k^5+1}$ looks like $\frac{k^3}{k^5} = \frac{1}{k^2}$. We will use the Limit Comparison Test to compare $\sum_{k=1}^{\infty} \frac{k^3-1}{k^5+1}$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$:

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{\frac{k^3-1}{k^5+1}}{\frac{1}{k^2}} &= \lim_{k \rightarrow \infty} \frac{k^5 - k^2}{k^5 + 1} \\ &= \lim_{k \rightarrow \infty} \frac{1 - \frac{1}{k^3}}{1 + \frac{1}{k^5}} \\ &= 1.\end{aligned}$$

Since this limit is 1, we know that the two series $\sum_{k=1}^{\infty} \frac{k^3-1}{k^5+1}$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$ either both converge or both diverge. Now $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is a p -series with $p = 2$, so $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges. Therefore, $\sum_{k=1}^{\infty} \frac{k^3-1}{k^5+1}$ converges as well.

- g. This series looks geometric, but to be sure let's write out the first few terms of this series:

$$\sum_{k=2}^{\infty} \frac{3^{k-1}}{7^k} = \frac{3}{7^2} + \frac{3^2}{7^3} + \frac{3^3}{7^4} + \cdots = \left(\frac{3}{7^2}\right) \left[1 + \frac{3}{7} + \left(\frac{3}{7}\right)^2 + \left(\frac{3}{7}\right)^3 + \cdots\right].$$

So $\sum_{k=2}^{\infty} \frac{3^{k-1}}{7^k}$ is a geometric series with $a = \frac{3}{49}$ and $r = \frac{3}{7}$. Since the ratio of this geometric series is between -1 and 1 , the series converges. The sum of the series is

$$\sum_{k=2}^{\infty} \frac{3^{k-1}}{7^k} = \left(\frac{3}{49}\right) \left(\frac{1}{1 - \frac{3}{7}}\right)$$

$$\begin{aligned}
 &= \left(\frac{3}{49}\right) \left(\frac{7}{4}\right) \\
 &= \frac{3}{28}.
 \end{aligned}$$

h. $\sum_{k=2}^{\infty} \frac{1}{k^k}$ converges by comparison with the series $\sum_{k=2}^{\infty} \frac{1}{k^2}$.

i. This series is an alternating series of the form $\sum_{k=1}^{\infty} (-1)^{k+1} a_k$ with $a_k = \frac{1}{\sqrt{k+1}}$. Since

$$\lim_{k \rightarrow \infty} \frac{1}{\sqrt{k+1}} = 0$$

and $\frac{1}{\sqrt{k+1}}$ decreases to 0, the Alternating Series Test shows that $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{\sqrt{k+1}}$ converges.

j. For this example we will use the Integral Test with the substitution $w = \ln(x)$, $dw = \frac{1}{x} dx$:

$$\begin{aligned}
 \int_2^{\infty} \frac{1}{x \ln(x)} dx &= \lim_{b \rightarrow \infty} \int_2^b \frac{1}{x \ln(x)} dx \\
 &= \lim_{b \rightarrow \infty} \int_{\ln(2)}^{\ln(b)} \frac{1}{w} dw \\
 &= \lim_{b \rightarrow \infty} \ln(w) \Big|_{\ln(2)}^{\ln(b)} \\
 &= \lim_{b \rightarrow \infty} [\ln(\ln(b)) - \ln(\ln(2))] \\
 &= \infty.
 \end{aligned}$$

Since $\int_2^{\infty} \frac{1}{x \ln(x)} dx$ diverges, we conclude $\sum_{k=2}^{\infty} \frac{1}{k \ln(k)}$ diverges.

k. First note that the terms $\frac{1}{\ln(k)}$ decrease to 0, so the Alternating Series Test shows that the alternating series $\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$ converges. Let $S = \sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$ and let S_n be the n th partial sum of this series. Recall that

$$|S_n - S| < |S_n - S_{n+1}| = a_{n+1}$$

where $a_{n+1} = \frac{1}{\ln(n+1)}$. If we can find a value of n so that $a_{n+1} < 0.001$, then $|S_n - S| < 0.001$ as desired. Now

$$\begin{aligned}
 a_{n+1} &< 0.001 \\
 \frac{1}{\ln(n+1)} &< 0.001 \\
 \ln(n+1) &> 1000 \\
 n+1 &> e^{1000} \\
 n &\geq e^{1000} - 1.
 \end{aligned}$$

So if we choose $n = e^{1000}$, then S_n approximates the sum of the series to within 0.001. Notice that $e^{1000} \approx 1.97 \times 10^{434}$ is a very large number. This shows that the series $\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$ converges very slowly.

8.4.5 Summary

- An alternating series is a series whose terms alternate in sign. It has the form

$$\sum (-1)^k a_k$$

where a_k is a positive real number for each k .

- The sequence of partial sums of a convergent alternating series oscillates around the sum of the series if the sequence of n th terms converges to 0. That is why the Alternating Series Test shows that the alternating series $\sum_{k=1}^{\infty} (-1)^k a_k$ converges whenever the sequence $\{a_n\}$ of n th terms decreases to 0.
- The difference between the $n - 1$ st partial sum S_{n-1} and the n th partial sum S_n of a convergent alternating series $\sum_{k=1}^{\infty} (-1)^k a_k$ is $|S_n - S_{n-1}| = a_n$. Since the partial sums oscillate around the sum S of the series, it follows that

$$|S - S_n| < a_n.$$

So the n th partial sum of a convergent alternating series $\sum_{k=1}^{\infty} (-1)^k a_k$ approximates the actual sum of the series to within a_n .

8.4.6 Exercises

1. (Testing convergence for an alternating series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Estimating the sum of an alternating series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Estimating the sum of a different alternating series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Estimating the sum of one more alternating series) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. Conditionally convergent series converge very slowly. As an example, consider the famous formula¹

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots = \sum_{k=0}^{\infty} (-1)^k \frac{1}{2k+1}. \quad (8.4.3)$$

In theory, the partial sums of this series could be used to approximate π .

- Show that the series in (8.4.3) converges conditionally.
- Let S_n be the n th partial sum of the series in (8.4.3). Calculate the error in approximating $\frac{\pi}{4}$ with S_{100} and explain why this is not a very good approximation.
- Determine the number of terms it would take in the series (8.4.3) to approximate $\frac{\pi}{4}$ to 10 decimal places. (The fact that it takes such a large number of terms to obtain even a modest degree of accuracy is why we say that conditionally convergent series converge very slowly.)

¹We will derive this formula in upcoming work.

Answer.

- a. $\sum_{k=0}^{\infty} \frac{1}{2k+1}$ diverges by comparison to the Harmonic series.
- b. $|S_{100} - \sum_{k=0}^{\infty} (-1)^k \frac{1}{2k+1}| < \approx 0.0049$.
- c. $n > 4,999,999,999.5$

Solution.

- a. Since $\sum_{k=0}^{\infty} (-1)^k \frac{1}{2k+1}$ is an alternating series and $\frac{1}{2k+1} \rightarrow 0$, the series converges by the Alternating Series Test. However, the corresponding series of positive terms $\sum_{k=0}^{\infty} \frac{1}{2k+1}$ diverges by comparison to the Harmonic series.

b.

$$|S_{100} - \sum_{k=0}^{\infty} (-1)^k \frac{1}{2k+1}| < a_{101} = \frac{1}{2 \cdot 101 + 1} \approx 0.0049$$

is a considerable amount of error when trying to estimate $\frac{\pi}{4}$ accurately.

- c. In this case we have

$$|S_n - \sum_{k=0}^{\infty} (-1)^k \frac{1}{2k+1}| < a_{n+1} = \frac{1}{2n+1}.$$

To have $a_{n+1} < 10^{-10}$ would require

$$\begin{aligned} \frac{1}{2n+1} &< 10^{-10} \\ 2n+1 &> 10^{10} \\ n &> \frac{10^{10}-1}{2} \\ n &> 4,999,999,999.5. \end{aligned}$$

- 6. We have shown that if $\sum (-1)^{k+1} a_k$ is a convergent alternating series, then the sum S of the series lies between any two consecutive partial sums S_n . This suggests that the average $\frac{S_n + S_{n+1}}{2}$ is a better approximation to S than is S_n .

- a. Show that $\frac{S_n + S_{n+1}}{2} = S_n + \frac{1}{2}(-1)^{n+2} a_{n+1}$.
- b. Use this revised approximation in (a) with $n = 20$ to approximate $\ln(2)$ given that

$$\ln(2) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k}.$$

Compare this to the approximation using just S_{20} . For your convenience, $S_{20} = \frac{155685007}{232792560}$.

Answer.

- a. $\frac{S_n + S_{n+1}}{2} = \frac{S_n + S_n + (-1)^{n+2} a_{n+1}}{2}.$
- b. $S_{20} = 0.668771403 \dots; \frac{S_{20} + S_{21}}{2} = \frac{161227687}{232792560} = 0.692580926 \dots$, accurate to within about 0.0006.

Solution.

- a. Since $S_{n+1} = S_n + (-1)^{n+2}a_{n+1}$ we have

$$\begin{aligned}\frac{S_n + S_{n+1}}{2} &= \frac{S_n + S_n + (-1)^{n+2}a_{n+1}}{2} \\ &= \frac{2S_n + (-1)^{n+2}a_{n+1}}{2} \\ &= S_n + \frac{1}{2}(-1)^{n+2}a_{n+1}.\end{aligned}$$

- b. We have two options for estimating $\ln(2)$ using partial sums. For the first one, we use

$$\ln(2) \approx S_{20} = \frac{155685007}{232792560} = 0.668771403\dots,$$

and note that $\ln(2) \approx 0.6931471$, so the error in approximating with S_{20} is about 0.025. If we instead use $\frac{S_{20}+S_{21}}{2}$, we get

$$\ln(2) \approx \frac{\frac{155685007}{232792560} + \left(\frac{155685007}{232792560} + \frac{1}{21}\right)}{2} = \frac{161227687}{232792560}$$

and since

$$\frac{161227687}{232792560} = 0.692580926\dots,$$

we see that we've obtained a much more accurate approximation of $\ln(2)$, accurate to within about 0.0006.

7. In this exercise, we examine one of the conditions of the Alternating Series Test. Consider the alternating series

$$1 - 1 + \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{9} + \frac{1}{4} - \frac{1}{16} + \dots,$$

where the terms are selected alternately from the sequences $\{\frac{1}{n}\}$ and $\{-\frac{1}{n^2}\}$.

- a. Explain why the n th term of the given series converges to 0 as n goes to infinity.
b. Rewrite the given series by grouping terms in the following manner:

$$(1 - 1) + \left(\frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{9}\right) + \left(\frac{1}{4} - \frac{1}{16}\right) + \dots.$$

Use this regrouping to determine if the series converges or diverges.

- c. Explain why the condition that the sequence $\{a_n\}$ *decreases* to a limit of 0 is included in the Alternating Series Test.

Answer.

- a. $\{\frac{1}{n}\}$ and $\{-\frac{1}{n^2}\}$ converge to 0.
b. Notice that $\frac{1}{k} - \frac{1}{k^2} = \frac{k-1}{k^2}$ and compare to the Harmonic series.
c. It is possible for a series to alternate, have the terms go to zero, have the terms not *decrease* to zero, and the series diverge.

Solution.

- a. Since both sequences $\left\{\frac{1}{n}\right\}$ and $\left\{-\frac{1}{n^2}\right\}$ converge to 0 as n goes to infinity, we see that the sequence of n th terms of our series converge to 0 as n goes to infinity.
- b. Notice that $\frac{1}{k} - \frac{1}{k^2} = \frac{k-1}{k^2}$. So our series can be rewritten as

$$\sum_{k=1}^{\infty} \frac{k-1}{k^2}.$$

Since

$$\lim_{k \rightarrow \infty} \frac{\frac{k-1}{k^2}}{\frac{1}{k}} = \lim_{k \rightarrow \infty} \frac{k^2 - k}{k^2} = 1,$$

the Limit Comparison Test shows that the series $\sum_{k=1}^{\infty} \frac{k-1}{k^2}$ behaves like the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$. Since the harmonic series diverges, it follows that the series $\sum_{k=1}^{\infty} \frac{k-1}{k^2}$ diverges as well.

- c. We have just demonstrated an example where the series alternates, the terms go to zero, but the terms don't *decrease* to zero, and the series diverges. This shows that the condition that the series decreases is necessary in the AST.
- 8. Conditionally convergent series exhibit interesting and unexpected behavior. In this exercise we examine the conditionally convergent alternating harmonic series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$ and discover that addition is not commutative for conditionally convergent series. We will also encounter Riemann's Theorem concerning rearrangements of conditionally convergent series. Before we begin, we remind ourselves that

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln(2),$$

a fact which will be verified in a later section.

- a. First we make a quick analysis of the positive and negative terms of the alternating harmonic series.
 - i. Show that the series $\sum_{k=1}^{\infty} \frac{1}{2k}$ diverges.
 - ii. Show that the series $\sum_{k=1}^{\infty} \frac{1}{2k+1}$ diverges.
 - iii. Based on the results of the previous parts of this exercise, what can we say about the sums $\sum_{k=C}^{\infty} \frac{1}{2k}$ and $\sum_{k=C}^{\infty} \frac{1}{2k+1}$ for any positive integer C ? Be specific in your explanation.
- b. Recall addition of real numbers is commutative; that is

$$a + b = b + a$$

for any real numbers a and b . This property is valid for any sum of finitely many terms, but does this property extend when we add infinitely many terms together?

The answer is no, and something even more odd happens. Riemann's Theorem (after the nineteenth-century mathematician Georg Friedrich Bernhard Riemann) states that a conditionally convergent series can be rearranged to converge to *any* prescribed sum. More specifically, this means that if we choose any real number S , we can rearrange the

terms of the alternating harmonic series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$ so that the sum is S . To understand how Riemann's Theorem works, let's assume for the moment that the number S we want our rearrangement to converge to is positive. Our job is to find a way to order the sum of terms of the alternating harmonic series to converge to S .

- i. Explain how we know that, regardless of the value of S , we can find a partial sum P_1

$$P_1 = \sum_{k=1}^{n_1} \frac{1}{2k+1} = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n_1+1}$$

of the positive terms of the alternating harmonic series that equals or exceeds S . Let

$$S_1 = P_1.$$

- ii. Explain how we know that, regardless of the value of S_1 , we can find a partial sum N_1

$$N_1 = - \sum_{k=1}^{m_1} \frac{1}{2k} = -\frac{1}{2} - \frac{1}{4} - \frac{1}{6} - \cdots - \frac{1}{2m_1}$$

so that

$$S_2 = S_1 + N_1 \leq S.$$

- iii. Explain how we know that, regardless of the value of S_2 , we can find a partial sum P_2

$$P_2 = \sum_{k=n_1+1}^{n_2} \frac{1}{2k+1} = \frac{1}{2(n_1+1)+1} + \frac{1}{2(n_1+2)+1} + \cdots + \frac{1}{2n_2+1}$$

of the remaining positive terms of the alternating harmonic series so that

$$S_3 = S_2 + P_2 \geq S.$$

- iv. Explain how we know that, regardless of the value of S_3 , we can find a partial sum

$$N_2 = - \sum_{k=m_1+1}^{m_2} \frac{1}{2k} = -\frac{1}{2(m_1+1)} - \frac{1}{2(m_1+2)} - \cdots - \frac{1}{2m_2}$$

of the remaining negative terms of the alternating harmonic series so that

$$S_4 = S_3 + N_2 \leq S.$$

- v. Explain why we can continue this process indefinitely and find a sequence $\{S_n\}$ whose terms are partial sums of a rearrangement of the terms in the alternating harmonic series so that $\lim_{n \rightarrow \infty} S_n = S$.

8.5 Taylor Polynomials and Taylor Series

Motivating Questions

- What is a Taylor polynomial? For what purposes are Taylor polynomials used?
- What is a Taylor series?
- How do we determine the accuracy when we use a Taylor polynomial to approximate a function?

So far, each infinite series we have discussed has been a series of real numbers, such as

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^k} + \cdots = \sum_{k=0}^{\infty} \frac{1}{2^k}. \quad (8.5.1)$$

In the remainder of this chapter, we will include series that involve a variable. For instance, if in the geometric series in Equation (8.5.1) we replace the ratio $r = \frac{1}{2}$ with the variable x , we have the infinite (still geometric) series

$$1 + x + x^2 + \cdots + x^k + \cdots = \sum_{k=0}^{\infty} x^k. \quad (8.5.2)$$

Here we see something very interesting: because a geometric series converges whenever its ratio r satisfies $|r| < 1$, and the sum of a convergent geometric series is $\frac{a}{1-r}$, we can say that for $|x| < 1$,

$$1 + x + x^2 + \cdots + x^k + \cdots = \frac{1}{1-x}. \quad (8.5.3)$$

Equation (8.5.3) states that the non-polynomial function $\frac{1}{1-x}$ on the right is equal to the infinite polynomial expression on the left. Because the terms on the left get very small as k gets large, we can truncate the series and say, for example, that

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

for small values of x . This shows one way that a polynomial function can be used to approximate a non-polynomial function; such approximations are one of the main themes in this section and the next.

In Example 8.5.1, we begin our exploration of approximating functions with polynomials.

Example 8.5.1. Example 8.3.1 showed how we can approximate the number e using linear, quadratic, and other polynomial functions; we then used similar ideas in Example 8.4.1 to approximate $\ln(2)$. In this example, we review and extend the process to find the “best” quadratic approximation to the exponential function e^x around the origin. Let $f(x) = e^x$ throughout this example.

- a. Find a formula for $P_1(x)$, the linearization of $f(x)$ at $x = 0$. (We label this linearization P_1 because it is a first degree polynomial approximation.) Recall that $P_1(x)$ is a good approximation to $f(x)$ for values of x close to 0. Plot f and P_1 near $x = 0$ to illustrate this fact.
- b. Since $f(x) = e^x$ is not linear, the linear approximation eventually is not a very good one. To obtain better approximations, we want to develop a different approximation that “bends” to make it more closely fit the graph of f near $x = 0$. To do so, we add a quadratic term to $P_1(x)$. In other words, we let

$$P_2(x) = P_1(x) + c_2x^2$$

for some real number c_2 . We need to determine the value of c_2 that makes the graph of $P_2(x)$ best fit the graph of $f(x)$ near $x = 0$.

Remember that $P_1(x)$ was a good linear approximation to $f(x)$ near 0; this is because $P_1(0) = f(0)$ and $P_1'(0) = f'(0)$. It is therefore reasonable to seek a value of c_2 so that

$$P_2(0) = f(0), \quad P_2'(0) = f'(0), \quad \text{and } P_2''(0) = f''(0).$$

Remember, we are letting $P_2(x) = P_1(x) + c_2x^2$.

- i. Calculate $P_2(0)$ to show that $P_2(0) = f(0)$.
- ii. Calculate $P_2'(0)$ to show that $P_2'(0) = f'(0)$.
- iii. Calculate $P_2''(x)$. Then find a value for c_2 so that $P_2''(0) = f''(0)$.
- iv. Explain why the condition $P_2''(0) = f''(0)$ will put an appropriate “bend” in the graph of P_2 to make P_2 fit the graph of f around $x = 0$.

Solution.

- a. We know that

$$P_1(x) = f(0) + f'(0)x = 1 + x.$$

Since $P_1(0) = f(0) = 1$ and $P_1'(0) = f'(0) = 1$, the graphs of P_1 and f agree at $x = 0$ and have the same slope at $x = 0$ (which means they go in the same direction at $x = 0$). This is why $P_1(x)$ is a good approximation to $f(x)$ for values of x close to 0.

- b. i. Since

$$P_2(x) = P_1(x) + c_2x^2 = f(0) + f'(0)x + c_2x^2$$

we have that

$$P_2(0) = 1 = f(0)$$

as desired.

- ii. A simple calculation shows $P_2'(x) = P_1'(x) + 2c_2x$. So $P_2'(0) = P_1'(0) = 1 = f'(0)$ as desired.
- iii. A simple calculation shows $P_2''(x) = 2c_2$. So $P_2''(0) = 2c_2$. To have $P_2''(0) = f''(0)$ we must have $2c_2 = f''(0)$ or $c_2 = \frac{f''(0)}{2} = \frac{1}{2}$.
- iv. The second derivative of a function tells us the concavity of the function. Concavity measures how the slopes of the tangent lines to the graph of the function are changing. This tells us how much bend there is in the graph. So if $P_2''(0) = f''(0)$, then P_2 will have the same bend in it at $x = 0$ as f does. This will make the graph of P_2 mold to the graph of f around $x = 0$.

8.5.1 Taylor Polynomials

Example 8.5.1 illustrates the first steps in the process of approximating functions with polynomials. Using this process we can approximate trigonometric, exponential, logarithmic, and other nonpolynomial functions as closely as we like (for certain values of x) with polynomials. This is extraordinarily useful in that it allows us to calculate values of these functions to whatever precision we like using only the operations of addition, subtraction, multiplication, and division, which can be easily programmed in a computer.

We next extend the approach in Example 8.5.1 to arbitrary functions at arbitrary points. Let f be a function that has as many derivatives as we need at a point $x = a$. Recall that $P_1(x)$ is the tangent line to f at $(a, f(a))$ and is given by the formula

$$P_1(x) = f(a) + f'(a)(x - a).$$

$P_1(x)$ is the linear approximation to f near a that has the same slope and function value as f at the point $x = a$.

We next want to find a quadratic approximation

$$P_2(x) = P_1(x) + c_2(x - a)^2$$

so that $P_2(x)$ more closely models $f(x)$ near $x = a$. Consider the following calculations of the values and derivatives of $P_2(x)$:

$$\begin{array}{ll} P_2(x) = P_1(x) + c_2(x - a)^2 & P_2(a) = P_1(a) = f(a) \\ P'_2(x) = P'_1(x) + 2c_2(x - a) & P'_2(a) = P'_1(a) = f'(a) \\ P''_2(x) = 2c_2 & P''_2(a) = 2c_2. \end{array}$$

To make $P_2(x)$ fit $f(x)$ better than $P_1(x)$, we want $P_2(x)$ and $f(x)$ to have the same concavity at $x = a$, in addition to having the same slope and function value. That is, we want to have

$$P''_2(a) = f''(a).$$

This implies that

$$2c_2 = f''(a)$$

and thus

$$c_2 = \frac{f''(a)}{2}.$$

Therefore, the quadratic approximation $P_2(x)$ to f centered at $x = a$ is

$$P_2(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2.$$

This approach extends naturally to polynomials of higher degree. We define polynomials

$$P_3(x) = P_2(x) + c_3(x - a)^3,$$

$$P_4(x) = P_3(x) + c_4(x - a)^4,$$

$$P_5(x) = P_4(x) + c_5(x - a)^5,$$

and in general

$$P_n(x) = P_{n-1}(x) + c_n(x - a)^n.$$

The defining property of these polynomials is that for each n , $P_n(x)$ and all its first n derivatives must agree with those of f at $x = a$. In other words we require that

$$P_n^{(k)}(a) = f^{(k)}(a)$$

for all k from 0 to n .

To see the conditions under which this happens, suppose

$$P_n(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + \cdots + c_n(x - a)^n.$$

Then

$$\begin{aligned} P_n^{(0)}(a) &= c_0 \\ P_n^{(1)}(a) &= c_1 \\ P_n^{(2)}(a) &= 2c_2 \\ P_n^{(3)}(a) &= (2)(3)c_3 \\ P_n^{(4)}(a) &= (2)(3)(4)c_4 \\ P_n^{(5)}(a) &= (2)(3)(4)(5)c_5 \end{aligned}$$

and, in general,

$$P_n^{(k)}(a) = (2)(3)(4) \cdots (k-1)(k)c_k = k!c_k.$$

So having $P_n^{(k)}(a) = f^{(k)}(a)$ means that $k!c_k = f^{(k)}(a)$ and therefore

$$c_k = \frac{f^{(k)}(a)}{k!}$$

for each value of k . Using this expression for c_k , we have found the formula for the polynomial approximation of f that we seek. Such a polynomial is called a *Taylor polynomial*.

Taylor Polynomials

The n th order Taylor polynomial of f centered at $x = a$ is given by

$$\begin{aligned} P_n(x) &= f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n \\ &= \sum_{k=0}^n \frac{f^{(k)}(a)}{k!}(x - a)^k. \end{aligned}$$

This degree n polynomial approximates $f(x)$ near $x = a$ and has the property that $P_n^{(k)}(a) = f^{(k)}(a)$ for $k = 0, 1, \dots, n$.

Example 8.5.2. Determine the third order Taylor polynomial for $f(x) = e^x$, as well as the general n th order Taylor polynomial for f centered at $x = 0$.

Solution. We know that $f'(x) = e^x$ and so $f''(x) = e^x$ and $f'''(x) = e^x$. Thus,

$$f(0) = f'(0) = f''(0) = f'''(0) = 1.$$

So the third order Taylor polynomial of $f(x) = e^x$ centered at $x = 0$ is

$$\begin{aligned} P_3(x) &= f(0) + f'(0)(x - 0) + \frac{f''(0)}{2!}(x - 0)^2 + \frac{f'''(0)}{3!}(x - 0)^3 \\ &= 1 + x + \frac{x^2}{2} + \frac{x^3}{6}. \end{aligned}$$

In general, for the exponential function f we have $f^{(k)}(x) = e^x$ for every positive integer k . Thus, the k th term in the n th order Taylor polynomial for $f(x)$ centered at $x = 0$ is

$$\frac{f^{(k)}(0)}{k!}(x - 0)^k = \frac{1}{k!}x^k.$$

Therefore, the n th order Taylor polynomial for $f(x) = e^x$ centered at $x = 0$ is

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \dots + \frac{1}{n!}x^n = \sum_{k=0}^n \frac{x^k}{k!}.$$

Example 8.5.3. We have just seen that the n th order Taylor polynomial centered at $a = 0$ for the exponential function e^x is

$$\sum_{k=0}^n \frac{x^k}{k!}.$$

In this example, we determine small order Taylor polynomials for several other familiar functions, and look for general patterns.

- Let $f(x) = \frac{1}{1-x}$.
 - Calculate the first four derivatives of $f(x)$ at $x = 0$. Then find the fourth order Taylor polynomial $P_4(x)$ for $\frac{1}{1-x}$ centered at 0.
 - Based on your results from part (i), determine a general formula for $f^{(k)}(0)$.
- Let $f(x) = \cos(x)$.
 - Calculate the first four derivatives of $f(x)$ at $x = 0$. Then find the fourth order Taylor polynomial $P_4(x)$ for $\cos(x)$ centered at 0.

- ii. Based on your results from part (i), find a general formula for $f^{(k)}(0)$. (Think about how k being even or odd affects the value of the k th derivative.)
- c. Let $f(x) = \sin(x)$.
 - i. Calculate the first four derivatives of $f(x)$ at $x = 0$. Then find the fourth order Taylor polynomial $P_4(x)$ for $\sin(x)$ centered at 0.
 - ii. Based on your results from part (i), find a general formula for $f^{(k)}(0)$. (Think about how k being even or odd affects the value of the k th derivative.)

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a. i. $f^k(0) = 0$ if k is odd, and $f^{2k}(0) = (-1)^k$.
 ii. $P_n(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots + (-1)^{n/2} \frac{x^n}{n!}$ if n is even and $P_n(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots + (-1)^{(n-1)/2} \frac{x^{(n-1)}}{(n-1)!}$ if n is odd.
- b. i. $f^k(0) = 0$ if k is even and $f^{2k+1}(0) = (-1)^k$.
 ii. $P_n(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots + (-1)^{(n-1)/2} \frac{x^n}{n!}$ if n is odd and $P_n(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots + (-1)^{n/2+1} \frac{x^{n-1}}{(n-1)!}$ if n is even.
- c. i. $f^{(k)}(0) = k!$.
 ii.

$$P_n(x) = \sum_{k=0}^n x^k.$$

Solution.

- a. i. The first four derivatives of $f(x)$ at $x = 0$ are

$$\begin{array}{ll}
 f(x) = \cos(x) & f(0) = 1 \\
 f'(x) = -\sin(x) & f'(0) = 0 \\
 f''(x) = -\cos(x) & f''(0) = -1 \\
 f^{(3)}(x) = \sin(x) & f^{(3)}(0) = 0 \\
 f^{(4)}(x) = \cos(x) & f^{(4)}(0) = 1.
 \end{array}$$

It appears that the odd derivatives of $f(x)$ are all plus or minus $\sin(x)$ and so have values of 0 at $x = 0$ and the even derivatives are $\pm \cos(x)$ and have alternating values of 1 and -1 at $x = 0$. Since the even numbers can be represented in the form $2k$ where k is an integer we have

$$f^k(0) = 0 \text{ if } k \text{ is odd, and } f^{2k}(0) = (-1)^k.$$

- ii. Based on the previous part of this problem the n th order Taylor polynomial for $\cos(x)$ is

$$1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^{n/2} \frac{x^n}{n!}$$

if n is even and

$$1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^{(n-1)/2} \frac{x^{(n-1)}}{(n-1)!}$$

if n is odd.

- b. i. The first four derivatives of $f(x)$ at $x = 0$ are

$$\begin{array}{ll} f(x) = \sin(x) & f(0) = 0 \\ f'(x) = \cos(x) & f'(0) = 1 \\ f''(x) = -\sin(x) & f''(0) = 0 \\ f^{(3)}(x) = -\cos(x) & f^{(3)}(0) = -1 \\ f^{(4)}(x) = \sin(x) & f^{(4)}(0) = 0. \end{array}$$

It appears that the even derivatives of $f(x)$ are all plus or minus $\sin(x)$ and so have values of 0 at $x = 0$ and the odd derivatives are $\pm \cos(x)$ and have alternating values of 1 and -1 at $x = 0$. Since the odd numbers can be represented in the form $2k + 1$ where k is an integer we have

$$f^k(0) = 0 \text{ if } k \text{ is even} \quad \text{and} \quad f^{2k+1}(0) = (-1)^k.$$

- ii. Based on the previous part of this problem the n th order Taylor polynomial for $\sin(x)$ is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{(n-1)/2} \frac{x^n}{n!}$$

if n is odd and

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n/2+1} \frac{x^{n-1}}{(n-1)!}$$

if n is even.

- c. i. The first four derivatives of $f(x)$ at $x = 0$ are

$$\begin{array}{ll} f(x) = \frac{1}{1-x} & f(0) = 1 \\ f'(x) = \frac{1}{(1-x)^2} & f'(0) = 1 \\ f''(x) = \frac{2}{(1-x)^3} & f''(0) = 2 \\ f^{(3)}(x) = \frac{3!}{(1-x)^4} & f^{(3)}(0) = 3! \\ f^{(4)}(x) = \frac{4!}{(1-x)^5} & f^{(4)}(0) = 4!. \end{array}$$

It appears that the pattern is

$$f^{(k)}(0) = k!.$$

ii. The n th order Taylor polynomial for f at $x = 0$ is

$$\sum_{k=0}^n \frac{f^{(k)}}{k!} x^k = \sum_{k=0}^n \frac{k!}{k!} x^k = \sum_{k=0}^n x^k.$$

This makes sense since $f(x)$ is the sum of the geometric series with ratio x , so the n th order Taylor polynomial should just be the n th partial sum of this geometric series.

It is possible that an n th order Taylor polynomial is not a polynomial of degree n ; that is, the order of the approximation can be different from the degree of the polynomial. For example, in Example 8.5.6 we found that the second order Taylor polynomial $P_2(x)$ centered at 0 for $\sin(x)$ is $P_2(x) = x$. In this case, the second order Taylor polynomial is a degree 1 polynomial.

8.5.2 Taylor Series

In Example 8.5.3 we saw that the fourth order Taylor polynomial $P_4(x)$ for $\sin(x)$ centered at 0 is

$$P_4(x) = x - \frac{x^3}{3!}.$$

The pattern we found for the derivatives $f^{(k)}(0)$ describe the higher-order Taylor polynomials, e.g.,

$$\begin{aligned} P_5(x) &= x - \frac{x^3}{3!} + \frac{x^{(5)}}{5!}, \\ P_7(x) &= x - \frac{x^3}{3!} + \frac{x^{(5)}}{5!} - \frac{x^{(7)}}{7!}, \\ P_9(x) &= x - \frac{x^3}{3!} + \frac{x^{(5)}}{5!} - \frac{x^{(7)}}{7!} + \frac{x^{(9)}}{9!}, \end{aligned}$$

and so on. It is instructive to consider the graphical behavior of these functions; Figure 8.5.4 shows the graphs of a few of the Taylor polynomials centered at 0 for the sine function.

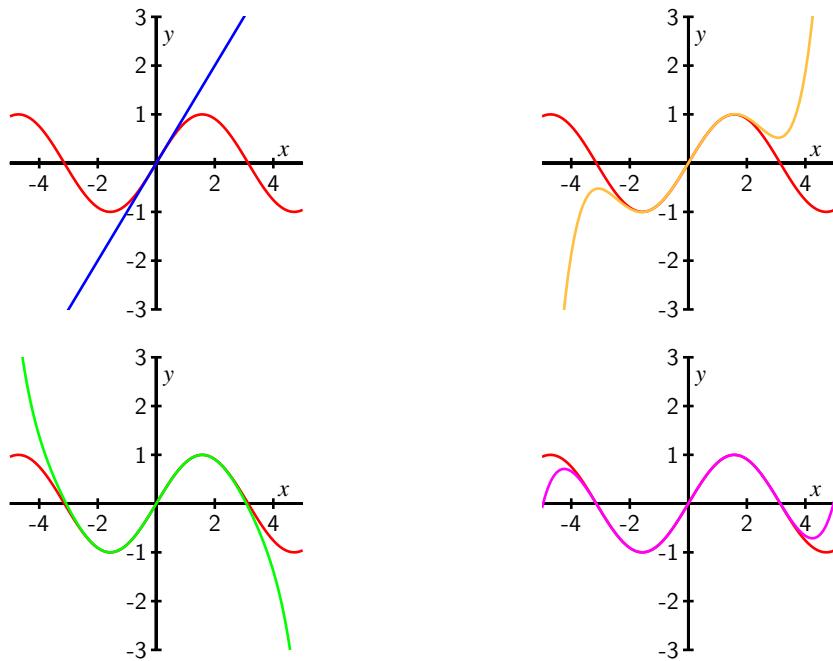


Figure 8.5.4: The order 1, 5, 7, and 9 Taylor polynomials centered at $x = 0$ for $f(x) = \sin(x)$.

Notice that $P_1(x)$ is close to the sine function only for values of x that are close to 0, but as we increase the degree of the Taylor polynomial the Taylor polynomials provide a better fit to the graph of the sine function over larger intervals. This illustrates the general behavior of Taylor polynomials: for any sufficiently well-behaved function, the sequence $\{P_n(x)\}$ of Taylor polynomials converges to the function f on larger and larger intervals (though those intervals may not necessarily increase without bound). If the Taylor polynomials ultimately converge to f on its entire domain, we write

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

Definition 8.5.5. Let f be a function all of whose derivatives exist at $x = a$. The **Taylor series** for f centered at $x = a$ is the series $T_f(x)$ defined by

$$T_f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x - a)^k.$$

In the special case where $a = 0$ in Definition 8.5.5, the Taylor series is also called the **Maclaurin series** for f . From Example 8.5.2 we know the n th order Taylor polynomial centered at 0 for the exponential function e^x ; thus, the Maclaurin series for e^x is

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Example 8.5.6. In Example 8.5.3 we determined small order Taylor polynomials for a few familiar functions, and also found general patterns in the derivatives evaluated at 0. Use that information to write the Taylor series centered at 0 for the following functions.

- a. $f(x) = \frac{1}{1-x}$
- b. $f(x) = \cos(x)$ (You will need to carefully consider how to indicate that many of the coefficients are 0. Think about a general way to represent an even integer.)
- c. $f(x) = \sin(x)$ (You will need to carefully consider how to indicate that many of the coefficients are 0. Think about a general way to represent an odd integer.)
- d. Determine the n order Taylor polynomial for $f(x) = \frac{1}{1-x}$ centered at $x = 0$.

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a. $P(x) = 1 + x + x^2 + x^3 + \cdots + x^n + \cdots$
- b. $P(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^n \frac{1}{(2n)!}x^{2n} + \cdots$
- c. $P(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + (-1)^n \frac{1}{(2n+1)!}x^{2n+1} + \cdots$
- d. $P_n(x) = 1 + x + x^2 + x^3 + \cdots + x^n$

Solution.

- a. For $f(x) = \frac{1}{1-x}$, its Taylor series is

$$P(x) = 1 + x + x^2 + x^3 + \cdots + x^n + \cdots$$

- b. For $f(x) = \cos(x)$, its Taylor series is

$$P(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^n \frac{1}{(2n)!}x^{2n} + \cdots$$

- c. For $f(x) = \sin(x)$, its Taylor series is

$$P(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + (-1)^n \frac{1}{(2n+1)!}x^{2n+1} + \cdots$$

- d. For $f(x) = \frac{1}{1-x}$,

$$P_n(x) = 1 + x + x^2 + x^3 + \cdots + x^n$$

Example 8.5.7.

- Plot the graphs of several of the Taylor polynomials centered at 0 (of order at least 5) for e^x and convince yourself that these Taylor polynomials converge to e^x for every value of x .
- Draw the graphs of several of the Taylor polynomials centered at 0 (of order at least 6) for $\cos(x)$ and convince yourself that these Taylor polynomials converge to $\cos(x)$ for every value of x . Write the Taylor series centered at 0 for $\cos(x)$.
- Draw the graphs of several of the Taylor polynomials centered at 0 for $\frac{1}{1-x}$. Based on your graphs, for what values of x do these Taylor polynomials appear to converge to $\frac{1}{1-x}$? How is this situation different from what we observe with e^x and $\cos(x)$? In addition, write the Taylor series centered at 0 for $\frac{1}{1-x}$.

Hint.

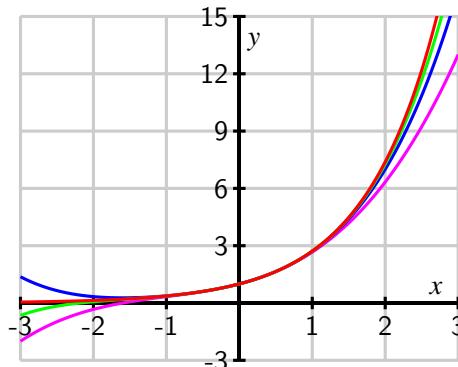
- Small hints for each of the prompts above.

Answer.

- It appears that as we increase the order of the Taylor polynomials, they fit the graph of f better and better over larger intervals.
- It appears that as we increase the order of the Taylor polynomials, they fit the graph of f better and better over larger intervals.
- The Taylor polynomials converge to $\frac{1}{1-x}$ only on the interval $(-1, 1)$.

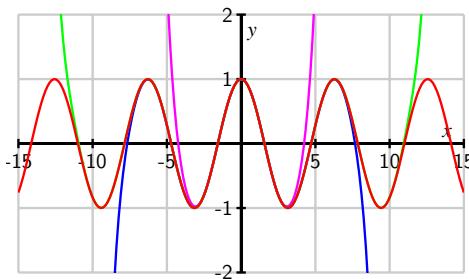
Solution.

- The graphs of the 10th (magenta), 20th (blue), and 30th (green) Taylor polynomials centered at 0 for e^x are shown below along with the graph of $f(x)$ in red:



It appears that as we increase the order of the Taylor polynomials, they fit the graph of f better and better over larger intervals. So it looks like the Taylor polynomials converge to e^x for every value of x .

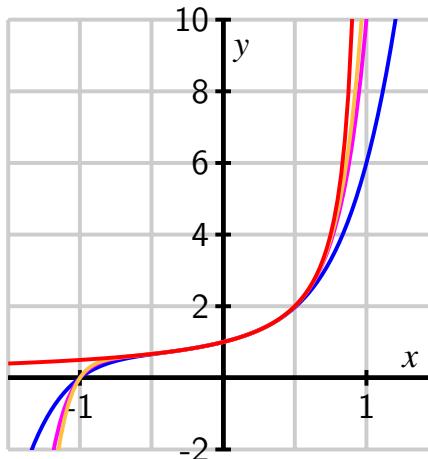
- b. The graphs of the 10th (magenta), 20th (blue), and 30th (green) Taylor polynomials centered at 0 for $\cos(x)$ are shown below along with the graph of $f(x)$ in red:



It appears that as we increase the order of the Taylor polynomials, they fit the graph of f better and better over larger intervals. So it looks like the Taylor polynomials converge to $\cos(x)$ for every value of x . Based on the n th order Taylor polynomials we found earlier for $\cos(x)$, the Taylor series for $f(x)$ centered at 0 is

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}.$$

- c. The graphs of the 10th (magenta), 20th (blue), and 30th (green) Taylor polynomials centered at 0 for $\frac{1}{1-x}$ are shown below along with the graph of $f(x)$ in red:



It appears that as we increase the order of the Taylor polynomials, they only fit the graph of f better and better over the interval $(-1, 1)$ and appear to diverge outside that interval. So it looks like the Taylor polynomials converge to $\frac{1}{1-x}$ only on the interval $(-1, 1)$.

Based on the n th order Taylor polynomials we found earlier for $\frac{1}{1-x}$, the Taylor series for $f(x)$ centered at 0 is

$$\sum_{k=0}^{\infty} x^k.$$

The Maclaurin series for e^x , $\sin(x)$, $\cos(x)$, and $\frac{1}{1-x}$ will be used frequently, so we should be certain to know and recognize them well.

8.5.3 The Interval of Convergence of a Taylor Series

In the previous section (in Figure 8.5.4 and Example 8.5.7) we observed that the Taylor polynomials centered at 0 for e^x , $\cos(x)$, and $\sin(x)$ converged to these functions for all values of x in their domain, but that the Taylor polynomials centered at 0 for $\frac{1}{1-x}$ converge to $\frac{1}{1-x}$ on the interval $(-1, 1)$ and diverge for all other values of x . So the Taylor series for a function $f(x)$ does not need to converge for all values of x in the domain of f .

Our observations suggest two natural questions: can we determine the values of x for which a given Taylor series converges? And does the Taylor series for a function f actually converge to $f(x)$?

Example 8.5.8. Graphical evidence suggests that the Taylor series centered at 0 for e^x converges for all values of x . To verify this, use the Ratio Test to determine all values of x for which the Taylor series

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (8.5.4)$$

converges absolutely.

Solution. Recall that the Ratio Test applies only to series of nonnegative terms. In this example, the variable x may have negative values. But we are interested in absolute convergence, so we apply the Ratio Test to the series

$$\sum_{k=0}^{\infty} \left| \frac{x^k}{k!} \right| = \sum_{k=0}^{\infty} \frac{|x|^k}{k!}.$$

Now, observe that

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} &= \lim_{k \rightarrow \infty} \frac{\frac{|x|^{k+1}}{(k+1)!}}{\frac{|x|^k}{k}} \\ &= \lim_{k \rightarrow \infty} \frac{|x|^{k+1} k!}{|x|^k (k+1)!} \\ &= \lim_{k \rightarrow \infty} \frac{|x|}{k+1} \\ &= 0 \end{aligned}$$

for any value of x . So the Taylor series (8.5.4) converges absolutely for every value of x , and thus converges for every value of x .

One question still remains: while the Taylor series for e^x converges for all x , what we have done does not tell us that this Taylor series actually converges to e^x for each x . We'll return

to this question when we consider the error in a Taylor approximation near the end of this section.

We can apply the main idea from Example 8.5.8 in general. To determine the values of x for which a Taylor series

$$\sum_{k=0}^{\infty} c_k(x-a)^k,$$

centered at $x = a$ will converge, we apply the Ratio Test with $a_k = |c_k(x-a)^k|$. The series converges if $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} < 1$.

Observe that

$$\frac{a_{k+1}}{a_k} = |x-a| \frac{|c_{k+1}|}{|c_k|},$$

so when we apply the Ratio Test, we get

$$\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = \lim_{k \rightarrow \infty} |x-a| \frac{|c_{k+1}|}{|c_k|}.$$

Note suppose that

$$\lim_{k \rightarrow \infty} \frac{|c_{k+1}|}{|c_k|} = L,$$

so that

$$\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_k} = |x-a| \cdot L.$$

There are three possibilities for L : L can be 0, it can be a finite positive value, or it can be infinite. Based on this value of L , we can determine for which values of x the original Taylor series converges.

- If $L = 0$, then the Taylor series converges on $(-\infty, \infty)$.
- If L is infinite, then the Taylor series converges only at $x = a$.
- If L is finite and nonzero, then the Taylor series converges absolutely for all x that satisfy

$$|x-a| \cdot L < 1$$

or for all x such that

$$|x-a| < \frac{1}{L},$$

which is the interval

$$\left(a - \frac{1}{L}, a + \frac{1}{L}\right).$$

Because the Ratio Test is inconclusive when the $|x-a| \cdot L = 1$, the endpoints $a \pm \frac{1}{L}$ have to be checked separately.

It is important to notice that the set of x values at which a Taylor series converges is always an interval centered at $x = a$. For this reason, the set on which a Taylor series converges is called the *interval of convergence*. Half the length of the interval of convergence is called the *radius of convergence*. If the interval of convergence of a Taylor series is infinite, then we say that the radius of convergence is infinite.

Example 8.5.9.

- Use the Ratio Test to explicitly determine the interval of convergence of the Taylor series for $f(x) = \frac{1}{1-x}$ centered at $x = 0$.
- Use the Ratio Test to explicitly determine the interval of convergence of the Taylor series for $f(x) = \cos(x)$ centered at $x = 0$.
- Use the Ratio Test to explicitly determine the interval of convergence of the Taylor series for $f(x) = \sin(x)$ centered at $x = 0$.

Hint.

- Small hints for each of the prompts above.

Answer.

- $(-\infty, \infty)$.
- $(-\infty, \infty)$.
- The interval $(-1, 1)$.

Solution.

- Using the Ratio Test with the k th term $\frac{|x|^{2k}}{(2k)!}$ we get

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{|x|^{2(k+1)}}{(2(k+1))!}}{\frac{|x|^{2k}}{(2k)!}} &= \lim_{k \rightarrow \infty} \frac{|x|^{2(k+1)}(2k)!}{|x|^{2k}(2(k+1))!} \\ &= \lim_{k \rightarrow \infty} \frac{|x|^2}{(2k+2)(2k+1)} \\ &= 0. \end{aligned}$$

So the interval of convergence of the Taylor series for $f(x) = \cos(x)$ centered at $x = 0$ is $(-\infty, \infty)$.

- Using the Ratio Test with the k th term $\frac{|x|^{2k+1}}{(2k+1)!}$ we get

$$\lim_{k \rightarrow \infty} \frac{\frac{|x|^{2(k+1)+1}}{(2(k+1)+1)!}}{\frac{|x|^{2k+1}}{(2k+1)!}} = \lim_{k \rightarrow \infty} \frac{|x|^{2(k+1)+1}(2k+1)!}{|x|^{2k+1}(2(k+1)+1)!}$$

$$\begin{aligned}
 &= \lim_{k \rightarrow \infty} \frac{|x|^2}{(2k+3)(2k+2)} \\
 &= 0.
 \end{aligned}$$

So the interval of convergence of the Taylor series for $f(x) = \sin(x)$ centered at $x = 0$ is $(-\infty, \infty)$.

- c. Using the Ratio Test with the k th term $|x|^k$ we get

$$\lim_{k \rightarrow \infty} \frac{|x|^{k+1}}{|x|^k} = \lim_{k \rightarrow \infty} |x| = |x|,$$

So the series $\sum_{k=0}^{\infty} x^k$ converges absolutely when $|x| < 1$ or for $-1 < x < 1$ and diverges when $|x| > 1$. Since the Ratio Test doesn't tell us what happens when $x = 1$, we need to check the endpoints separately.

- When $x = 1$ we have the series $\sum_{k=0}^{\infty} 1$ which diverges since $\lim_{k \rightarrow \infty} 1 \neq 0$.
- When $x = -1$ we have the series $\sum_{k=0}^{\infty} (-1)^k$ which diverges since $\lim_{k \rightarrow \infty} (-1)^k$ does not exist.

Therefore, the interval of convergence of the Taylor series for $f(x) = \frac{1}{1-x}$ centered at $x = 0$ is $(-1, 1)$.

The Ratio Test allows us to determine the set of x values for which a Taylor series converges absolutely. However, just because a Taylor series for a function f converges, we cannot be certain that the Taylor series actually converges to $f(x)$. To show why and where a Taylor series does in fact converge to the function f , we next consider the error that is present in Taylor polynomials.

8.5.4 Error Approximations for Taylor Polynomials

We now know how to find Taylor polynomials for functions such as $\sin(x)$, and how to determine the interval of convergence of the corresponding Taylor series. We next develop an error bound that will tell us how well an n th order Taylor polynomial $P_n(x)$ approximates its generating function $f(x)$. This error bound will also allow us to determine whether a Taylor series on its interval of convergence actually equals the function f from which the Taylor series is derived. Finally, we will be able to use the error bound to determine the order of the Taylor polynomial $P_n(x)$ that we will ensure that $P_n(x)$ approximates $f(x)$ to the desired degree of accuracy.

For this argument, we assume throughout that we center our approximations at 0 (but a similar argument holds for approximations centered at a). We define the exact error, $E_n(x)$, that results from approximating $f(x)$ with $P_n(x)$ by

$$E_n(x) = f(x) - P_n(x).$$

We are particularly interested in $|E_n(x)|$, the distance between P_n and f . Because

$$P_n^{(k)}(0) = f^{(k)}(0)$$

for $0 \leq k \leq n$, we know that

$$E_n^{(k)}(0) = 0$$

for $0 \leq k \leq n$. Furthermore, since $P_n(x)$ is a polynomial of degree less than or equal to n , we know that

$$P_n^{(n+1)}(x) = 0.$$

Thus, since $E_n^{(n+1)}(x) = f^{(n+1)}(x) - P_n^{(n+1)}(x)$, it follows that

$$E_n^{(n+1)}(x) = f^{(n+1)}(x)$$

for all x .

Suppose that we want to approximate $f(x)$ at a number c close to 0 using $P_n(c)$. If we assume $|f^{(n+1)}(t)|$ is bounded by some number M on $[0, c]$, so that

$$|f^{(n+1)}(t)| \leq M$$

for all $0 \leq t \leq c$, then we can say that

$$|E_n^{(n+1)}(t)| = |f^{(n+1)}(t)| \leq M$$

for all t between 0 and c . Equivalently,

$$-M \leq E_n^{(n+1)}(t) \leq M \quad (8.5.5)$$

on $[0, c]$. Next, we integrate the three terms in Inequality (8.5.5) from $t = 0$ to $t = x$, and thus find that

$$\int_0^x -M \, dt \leq \int_0^x E_n^{(n+1)}(t) \, dt \leq \int_0^x M \, dt$$

for every value of x in $[0, c]$. Since $E_n^{(n)}(0) = 0$, the First FTC tells us that

$$-Mx \leq E_n^{(n)}(x) \leq Mx$$

for every x in $[0, c]$.

Integrating this last inequality, we obtain

$$\int_0^x -Mt \, dt \leq \int_0^x E_n^{(n)}(t) \, dt \leq \int_0^x Mt \, dt$$

and thus

$$-M \frac{x^2}{2} \leq E_n^{(n-1)}(x) \leq M \frac{x^2}{2}$$

for all x in $[0, c]$.

Integrating n times, we arrive at

$$-M \frac{x^{n+1}}{(n+1)!} \leq E_n(x) \leq M \frac{x^{n+1}}{(n+1)!}$$

for all x in $[0, c]$. This enables us to conclude that

$$|E_n(x)| \leq M \frac{|x|^{n+1}}{(n+1)!}$$

for all x in $[0, c]$, and we have found a bound on the approximation's error, E_n .

Our work above was based on the approximation centered at $a = 0$; the argument may be generalized to hold for any value of a , which results in the following theorem.

The Lagrange Error Bound for $P_n(x)$

Let f be a continuous function with $n+1$ continuous derivatives. Suppose that M is a positive real number such that $|f^{(n+1)}(x)| \leq M$ on the interval $[a, c]$. If $P_n(x)$ is the n th order Taylor polynomial for $f(x)$ centered at $x = a$, then

$$|P_n(c) - f(c)| \leq M \frac{|c - a|^{n+1}}{(n+1)!}.$$

We can use this error bound to tell us important information about Taylor polynomials and Taylor series, as we see in the following examples and activities.

Example 8.5.10. Determine how well the 10th order Taylor polynomial $P_{10}(x)$ for $\sin(x)$, centered at 0, approximates $\sin(2)$.

Solution. To answer this question we use $f(x) = \sin(x)$, $c = 2$, $a = 0$, and $n = 10$ in the Lagrange error bound formula. We also need to find an appropriate value for M . Note that the derivatives of $f(x) = \sin(x)$ are all equal to $\pm \sin(x)$ or $\pm \cos(x)$. Thus,

$$|f^{(n+1)}(x)| \leq 1$$

for any n and x . Therefore, we can choose M to be 1. Then

$$|P_{10}(2) - f(2)| \leq (1) \frac{|2 - 0|^{11}}{(11)!} = \frac{2^{11}}{(11)!} \approx 0.00005130671797.$$

So $P_{10}(2)$ approximates $\sin(2)$ to within at most 0.00005130671797. A computer algebra system tells us that

$$P_{10}(2) \approx 0.9093474427 \quad \text{and} \quad \sin(2) \approx 0.9092974268$$

with an actual difference of about 0.0000500159.

Example 8.5.11. Let $P_n(x)$ be the n th order Taylor polynomial for $\sin(x)$ centered at $x = 0$. Determine how large we need to choose n so that $P_n(2)$ approximates $\sin(2)$ to 20 decimal places.

Hint. Small hints for each of the prompts above.

Answer. $n \geq 27$.

Solution. In this example, if we can find a value of n so that

$$M \frac{|2 - 0|^{n+1}}{(n+1)!} < 10^{-20}$$

then we will have

$$|P_n(2) - f(2)| \leq M \frac{|2 - 0|^{n+1}}{(n+1)!} < 10^{-20}.$$

Again we use $f(x) = \sin(x)$, $c = 2$, $a = 0$, and $M = 1$ from the previous example. So we need to find n to make

$$\frac{2^{n+1}}{(n+1)!} \leq 10^{-20}.$$

There is no good way to solve equations involving factorials, so we simply use trial and error, evaluating $\frac{2^{n+1}}{(n+1)!}$ at different values of n until we get one we need.

n	$\frac{2^{n+1}}{(n+1)!}$
10	$5.130671797 \times 10^{-5}$
20	$4.104743250 \times 10^{-14}$
25	$1.664028884 \times 10^{-19}$
26	$1.232613988 \times 10^{-20}$
27	$8.804385630 \times 10^{-22}$

So we need to use an n of at least 27 to ensure accuracy to 20 decimal places.

A computer algebra system gives

$$\begin{aligned} P_{27}(2) &\approx 0.9092974268256816953960 \\ \sin(2) &\approx 0.9092974268256816953960 \end{aligned}$$

and we can see that these agree to 20 places.

Example 8.5.12. Show that the Taylor series for $\sin(x)$ actually converges to $\sin(x)$ for all x .

Solution. Recall from the previous example that since $f(x) = \sin(x)$, we know

$$|f^{(n+1)}(x)| \leq 1$$

for any n and x . This allows us to choose $M = 1$ in the Lagrange error bound formula. Thus,

$$|P_n(x) - \sin(x)| \leq \frac{|x|^{n+1}}{(n+1)!} \tag{8.5.6}$$

for every x .

We showed in earlier work that the Taylor series $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ converges for every value of x . Because the terms of any convergent series must approach zero, it follows that

$$\lim_{n \rightarrow \infty} \frac{x^{n+1}}{(n+1)!} = 0$$

for every value of x . Thus, taking the limit as $n \rightarrow \infty$ in the inequality (8.5.6), it follows that

$$\lim_{n \rightarrow \infty} |P_n(x) - \sin(x)| = 0.$$

As a result, we can now write

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

for every real number x .

Example 8.5.13.

- a. Show that the Taylor series centered at 0 for $\cos(x)$ converges to $\cos(x)$ for every real number x .
- b. Next we consider the Taylor series for e^x .
 - i. Show that the Taylor series centered at 0 for e^x converges to e^x for every nonnegative value of x .
 - ii. Show that the Taylor series centered at 0 for e^x converges to e^x for every negative value of x .
 - iii. Explain why the Taylor series centered at 0 for e^x converges to e^x for every real number x . Recall that we earlier showed that the Taylor series centered at 0 for e^x converges for all x , and we have now completed the argument that the Taylor series for e^x actually converges to e^x for all x .
- c. Let $P_n(x)$ be the n th order Taylor polynomial for e^x centered at 0. Find a value of n so that $P_n(5)$ approximates e^5 correct to 8 decimal places.

Hint.

- a. Small hints for each of the prompts above.

Answer.

- a. Compare Example 8.5.12.
- b.
 - i. Use the fact that $|f^{(n)}(x)| \leq e^c$ on the interval $[0, c]$ for any fixed positive value of c .
 - ii. Repeat the argument in (a) but replace e^c with 1, and everything else holds in the same way.
 - iii. Combine the results of (a) and (b)
- c. $n = 28$.

Solution.

- a. Compare Example 8.5.12.
- b. i. Let $x \geq 0$. Since $f(x) = e^x$, $f^{(n)}(x) = e^x$ for every natural number n . Since e^x is an increasing function, we know that $|f^{(n)}(x)| \leq e^c$ on the interval $[0, c]$ for any fixed positive value of c . Thus, by the Lagrange error formula, we can say that

$$|P_n(x) - e^x| \leq e^c \frac{x^{n+1}}{(n+1)!}.$$

Since the series $\sum \frac{x^n}{n!}$ converges for every x , $\frac{x^n}{n!} \rightarrow 0$ as $x \rightarrow \infty$, and thus $\frac{x^{n+1}}{(n+1)!} \rightarrow 0$ as $n \rightarrow \infty$ for every x in $[0, c]$. Further, since e^c is a constant independent of n , $e^c \frac{x^{n+1}}{(n+1)!} \rightarrow 0$ as well. Thus,

$$\lim_{n \rightarrow \infty} |P_n(x) - e^x| = 0,$$

as desired.

- ii. When $x < 0$, we know $e^x < 1$. Thus, we can repeat our argument in (a) but replace e^c with 1, and everything else holds in the same way.
- iii. Because we have shown that the Taylor series for e^x converges to e^x for both every nonnegative x -value and for every negative x -value, it follows that we have convergence for every value of x .
- c. Since e^x is increasing on $[0, 5]$ we know that $e^x < e^5$ on $[0, 5]$. Now $e^5 < 243$, so

$$|P_n(5) - e^5| \leq 243 \frac{|5|^{n+1}}{(n+1)!}.$$

We want a value of n that makes this error term less than 10^{-8} . Testing various values of n gives

$$243 \frac{|5|^{28+1}}{(28+1)!} \approx 5.119146745 \times 10^{-9}$$

so we can choose $n = 28$. A computer algebra system shows that $P_{28}(5) \approx 148.413159102551$ while $e^5 \approx 148.413159102577$ and we can see that these two approximations agree to 8 decimal places.

8.5.5 Summary

- We can use Taylor polynomials to approximate functions. This allows us to approximate values of functions using only addition, subtraction, multiplication, and division of real numbers. The n th order Taylor polynomial centered at $x = a$ of a function f is

$$\begin{aligned} P_n(x) &= f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n \\ &= \sum_{k=0}^n \frac{f^{(k)}(a)}{k!}(x - a)^k. \end{aligned}$$

- The Taylor series centered at $x = a$ for a function f is

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

The n th order Taylor polynomial centered at a for f is the n th partial sum of its Taylor series centered at a . So the n th order Taylor polynomial for a function f is an approximation to f on the interval where the Taylor series converges; for the values of x for which the Taylor series converges to f we write

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

- The Lagrange Error Bound shows us how to determine the accuracy in using a Taylor polynomial to approximate a function. More specifically, if $P_n(x)$ is the n th order Taylor polynomial for f centered at $x = a$ and if M is an upper bound for $|f^{(n+1)}(x)|$ on the interval $[a, c]$, then

$$|P_n(c) - f(c)| \leq M \frac{|c - a|^{n+1}}{(n + 1)!}.$$

8.5.6 Exercises

1. (Determining Taylor polynomials from a function formula) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Determining Taylor polynomials from given derivative values) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. (Finding the Taylor series for a given rational function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
4. (Finding the Taylor series for a given trigonometric function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
5. (Finding the Taylor series for a given logarithmic function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
6. In this exercise we investigate the Taylor series of polynomial functions.
 - a. Find the 3rd order Taylor polynomial centered at $a = 0$ for $f(x) = x^3 - 2x^2 + 3x - 1$. Does your answer surprise you? Explain.
 - b. Without doing any additional computation, find the 4th, 12th, and 100th order Taylor polynomials (centered at $a = 0$) for $f(x) = x^3 - 2x^2 + 3x - 1$. Why should you expect this?
 - c. Now suppose $f(x)$ is a degree m polynomial. Completely describe the n th order Taylor polynomial (centered at $a = 0$) for each n .

Answer.

- a. $P_3(x) = -1 + 3x - \frac{4}{2!}x^2 + \frac{6}{3!}x^3$, which is the same polynomial as $f(x)$.
- b. For $n \geq 3$, $P_n(x) = f(x)$.

- c. For $n \geq m$, $P_n(x) = f(x)$.

Solution.

- a. Observe that $f'(x) = 3x^2 - 4x + 3$, $f''(x) = 6x - 4$, and $f'''(x) = 6$, so $f'(0) = 3$, $f''(0) = -4$, and $f'''(0) = 6$. In addition, $f(0) = -1$. Thus, the 3rd order Taylor polynomial centered at $a = 0$ is

$$P_3(x) = -1 + 3x - \frac{4}{2!}x^2 + \frac{6}{3!}x^3.$$

Note that this is the same polynomial as $f(x)$, which is not surprising since the degree 3 Taylor polynomial is the degree 3 polynomial that has the same value, slope, concavity, and third derivative at $a = 0$ as the original.

- b. Since $f^{(4)}(x) = 0$ for all x , $f^{(4)}(0) = 0$, and similarly every higher derivative of f will be zero. Thus, the the 4th, 12th, and 100th order Taylor polynomials (centered at $a = 0$) are all identical to $P_3(x)$ found in (a). for $f(x) = x^3 - 2x^2 + 3x - 1$. We should expect this because all of the higher derivatives (than the third) of the original function are identically zero.
- c. For $n = 1, 2, \dots, m-1$, the degree n Taylor polynomial of f will be given by the usual formula. But for every $n \geq m$, $P_n(x) = f(x)$. That is, all of the Taylor polynomials of degree at least that of the original function are simply the function itself.

7. The examples we have considered in this section have all been for Taylor polynomials and series centered at 0, but Taylor polynomials and series can be centered at any value of a . We look at examples of such Taylor polynomials in this exercise.

- a. Let $f(x) = \sin(x)$. Find the Taylor polynomials up through order four of f centered at $x = \frac{\pi}{2}$. Then find the Taylor series for $f(x)$ centered at $x = \frac{\pi}{2}$. Why should you have expected the result?
- b. Let $f(x) = \ln(x)$. Find the Taylor polynomials up through order four of f centered at $x = 1$. Then find the Taylor series for $f(x)$ centered at $x = 1$.
- c. Use your result from (b) to determine which Taylor polynomial will approximate $\ln(2)$ to two decimal places. Explain in detail how you know you have the desired accuracy.

Answer.

a.

$$\begin{aligned} P_1(x) &= 1 + 0\left(x - \frac{\pi}{2}\right) = 1 \\ P_2(x) &= 1 + 0\left(x - \frac{\pi}{2}\right) - \frac{1}{2!}\left(x - \frac{\pi}{2}\right)^2 = 1 - \frac{1}{2!}\left(x - \frac{\pi}{2}\right)^2 \\ P_3(x) &= 1 + 0\left(x - \frac{\pi}{2}\right) - \frac{1}{2!}\left(x - \frac{\pi}{2}\right)^2 + \frac{0}{3!}\left(x - \frac{\pi}{2}\right)^3 \\ &= 1 - \frac{1}{2!}\left(x - \frac{\pi}{2}\right)^2 = P_2(x) \\ P_4(x) &= 1 + 0\left(x - \frac{\pi}{2}\right) - \frac{1}{2!}\left(x - \frac{\pi}{2}\right)^2 + \frac{0}{3!}\left(x - \frac{\pi}{2}\right)^3 + \frac{1}{4!}\left(x - \frac{\pi}{2}\right)^4 \\ &= 1 - \frac{1}{2!}\left(x - \frac{\pi}{2}\right)^2 + \frac{1}{4!}\left(x - \frac{\pi}{2}\right)^4 \end{aligned}$$

$$P(x) = 1 - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 + \frac{1}{4!} \left(x - \frac{\pi}{2} \right)^4 - \frac{1}{6!} \left(x - \frac{\pi}{2} \right)^6 + \dots$$

b.

$$P_4(x) = 0 + 1(x - 1) - \frac{1}{2!}(x - 1)^2 + \frac{2}{3!}(x - 1)^3 - \frac{6}{4!}(x - 1)^4.$$

$$P(x) = 1(x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 - \frac{1}{4}(x - 1)^4 + \frac{1}{5}(x - 1)^5 - \dots$$

c. $P_{101}(1) \approx 0.698073$ **Solution.**

- a. For $f(x) = \sin(x)$, $f'(x) = \cos(x)$, $f''(x) = -\sin(x)$, $f'''(x) = -\cos(x)$, and $f^{(4)}(x) = \sin(x)$. Thus, $f\left(\frac{\pi}{2}\right) = 1$, $f'f\left(\frac{\pi}{2}\right) = 0$, $f''f\left(\frac{\pi}{2}\right) = -1$, $f'''f\left(\frac{\pi}{2}\right) = 0$, and $f^{(4)}f\left(\frac{\pi}{2}\right) = 1$. It follows that the first four Taylor polynomials of f are

$$\begin{aligned} P_1(x) &= 1 + 0 \left(x - \frac{\pi}{2} \right) = 1 \\ P_2(x) &= 1 + 0 \left(x - \frac{\pi}{2} \right) - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 = 1 - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 \\ P_3(x) &= 1 + 0 \left(x - \frac{\pi}{2} \right) - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 + \frac{0}{3!} \left(x - \frac{\pi}{2} \right)^3 \\ &= 1 - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 = P_2(x) \\ P_4(x) &= 1 + 0 \left(x - \frac{\pi}{2} \right) - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 + \frac{0}{3!} \left(x - \frac{\pi}{2} \right)^3 + \frac{1}{4!} \left(x - \frac{\pi}{2} \right)^4 \\ &= 1 - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 + \frac{1}{4!} \left(x - \frac{\pi}{2} \right)^4 \end{aligned}$$

From the pattern, the Taylor series for $f(x)$ centered at $x = \frac{\pi}{2}$ is

$$P(x) = 1 - \frac{1}{2!} \left(x - \frac{\pi}{2} \right)^2 + \frac{1}{4!} \left(x - \frac{\pi}{2} \right)^4 - \frac{1}{6!} \left(x - \frac{\pi}{2} \right)^6 + \dots$$

which is expected because of the repeating patterns in the derivatives of the sine function evaluated at $\frac{\pi}{2}$.

- b. For $f(x) = \ln(x)$, $f'(x) = x^{-1}$, $f''(x) = -x^{-2}$, $f'''(x) = 2x^{-3}$, and $f^{(4)}(x) = -6x^{-4}$. It follows that $f(1) = 0$, $f'(1) = 1$, $f''(1) = -1$, $f'''(1) = 2$, and $f^{(4)}(1) = -6$. Thus, the fourth Taylor polynomial (in which we can see the polynomials of lower degree) is

$$P_4(x) = 0 + 1(x - 1) - \frac{1}{2!}(x - 1)^2 + \frac{2}{3!}(x - 1)^3 - \frac{6}{4!}(x - 1)^4.$$

Simplifying the coefficients and seeing the pattern, it follows that the Taylor series for $f(x)$ centered at $x = 1$ is

$$P(x) = 1(x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 - \frac{1}{4}(x - 1)^4 + \frac{1}{5}(x - 1)^5 - \dots$$

- c. To estimate $\ln(2)$ to two decimal places, we use the fact that $\ln(x) = 1(x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \frac{1}{5}(x-1)^5 - \dots$ for x near 1, and thus

$$\ln(2) = P(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

Because $P(2)$ is an alternating series, we need to know when the next term in the series is less than 0.01, which occurs when $n = 101$. Thus, if we compute $P_{101}(1) = \sum_{k=1}^{101} (-1)^{k+1} \frac{1}{k}$, we get the desired estimate. We note particularly that $P_{101}(1) \approx 0.698073$, which is indeed an estimate of $\ln(2)$ accurate to two decimal places.

8. We can use known Taylor series to obtain other Taylor series, and we explore that idea in this exercise, as a preview of work in the following section.

- a. Calculate the first four derivatives of $\sin(x^2)$ and hence find the fourth order Taylor polynomial for $\sin(x^2)$ centered at $a = 0$.
- b. Part (a) demonstrates the brute force approach to computing Taylor polynomials and series. Now we find an easier method that utilizes a known Taylor series. Recall that the Taylor series centered at 0 for $f(x) = \sin(x)$ is

$$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}. \quad (8.5.7)$$

- i. Substitute x^2 for x in the Taylor series (8.5.7). Write out the first several terms and compare to your work in part (a). Explain why the substitution in this problem should give the Taylor series for $\sin(x^2)$ centered at 0.
- ii. What should we expect the interval of convergence of the series for $\sin(x^2)$ to be? Explain in detail.

Answer.

- a. $P_4(x) = x^2$.
- b. i. $g(x) = P(x^2) = x^2 - \frac{1}{3!}x^6 + \frac{1}{5!}x^{10} - \dots$
- ii. All real numbers.

Solution.

- a. Let $f(x) = \sin(x^2)$. Then

$$\begin{array}{ll} \sin(x^2) & f(0) = 0 \\ f'(x) = 2x \cos(x^2) & f'(0) = 0 \\ f''(x) = -4x^2 \sin(x^2) + 2 \cos(x^2) & f''(0) = 0 \\ f'''(x) = -8x^3 \cos(x^2) - 12x \sin(x^2) & f'''(0) = 0 \\ f^{(4)}(x) = 16x^4 \sin(x^2) - 48x^2 \cos(x^2) - 12 \sin(x^2) & f^{(4)}(0) = 0 \end{array}$$

so the fourth order Taylor polynomial for f centered at $a = 0$ is

$$P_4(x) = x^2.$$

- b. i. Substituting x^2 for x in the Taylor series for $\sin(x)$, we find that since

$$P(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots,$$

we have

$$g(x) = P(x^2) = x^2 - \frac{1}{3!}x^6 + \frac{1}{5!}x^{10} - \dots$$

which should be the Taylor series for $\sin(x^2)$ since $\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$ for every value of x .

- ii. We expect the interval of convergence of the series for $\sin(x^2)$ to be the same as the series for $\sin(x)$, since $\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$ for every value of x .

9. Based on the examples we have seen, we might expect that the Taylor series for a function f always converges to the values $f(x)$ on its interval of convergence. We explore that idea

in more detail in this exercise. Let $f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$

- Show, using the definition of the derivative, that $f'(0) = 0$.
- It can be shown that $f^{(n)}(0) = 0$ for all $n \geq 2$. Assuming that this is true, find the Taylor series for f centered at 0.
- What is the interval of convergence of the Taylor series centered at 0 for f ? Explain. For which values of x the interval of convergence of the Taylor series does the Taylor series converge to $f(x)$?

8.6 Power Series

Motivating Questions

- What is a power series?
- What are some important uses of power series?
- What is the connection between power series and Taylor series?

We have noted in our work with Taylor polynomials and Taylor series that polynomial functions are the simplest possible functions in mathematics, in part because they require only addition and multiplication to evaluate. From the point of view of calculus, polynomials are especially nice: we can easily differentiate or integrate any polynomial. In light of our work in Section 8.5, we now know that several important non-polynomials have polynomial-like expansions. For example, for any real number x ,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$$

There are two settings where other series like the one for e^x arise: we may be given an expression such as

$$1 + 2x + 3x^2 + 4x^3 + \cdots$$

and we seek the values of x for which the expression makes sense. Or we may be trying to find an unknown function f that has expression

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k + \cdots,$$

and we try to determine the values of the constants a_0, a_1, \dots . The latter situation is explored in Example 8.6.1.

Example 8.6.1. In Chapter 7, we learned some of the many important applications of differential equations, and learned some approaches to solve or analyze them. Here, we consider an important approach that will allow us to solve a wider variety of differential equations.

Let's consider the familiar differential equation from exponential population growth given by

$$y' = ky, \tag{8.6.1}$$

where k is the constant of proportionality. While we can solve this differential equation using methods we have already learned, we take a different approach now that can be applied to a much larger set of differential equations. For the rest of this example, let's assume that $k = 1$. We will use our knowledge of Taylor series to find a solution to the differential equation (8.6.1).

To do so, we assume that we have a solution $y = f(x)$ and that $f(x)$ has a Taylor series that can be written in the form

$$y = f(x) = \sum_{k=0}^{\infty} a_k x^k,$$

where the coefficients a_k are undetermined. Our task is to find the coefficients.

- a. Assume that we can differentiate a power series term by term. By taking the derivative of $f(x)$ with respect to x and substituting the result into the differential equation (8.6.1), show that the equation

$$\sum_{k=1}^{\infty} ka_k x^{k-1} = \sum_{k=0}^{\infty} a_k x^k$$

must be satisfied in order for $f(x) = \sum_{k=0}^{\infty} a_k x^k$ to be a solution of the DE.

- b. Two series are equal if and only if they have the same coefficients on like power terms. Use this fact to find a relationship between a_1 and a_0 .
- c. Now write a_2 in terms of a_1 . Then write a_2 in terms of a_0 .
- d. Write a_3 in terms of a_2 . Then write a_3 in terms of a_0 .
- e. Write a_4 in terms of a_3 . Then write a_4 in terms of a_0 .
- f. Observe that there is a pattern in (b)-(e). Find a general formula for a_k in terms of a_0 .
- g. Write the series expansion for y using only the unknown coefficient a_0 . From this, determine what familiar functions satisfy the differential equation (8.6.1). (Hint: Compare to a familiar Taylor series.)

Solution.

- a. Differentiation term by term gives

$$y' = \sum_{k=1}^{\infty} ka_k x^{k-1}.$$

We then substitute this series into the differential equation (8.6.1) to obtain the equation

- b. When we write the first few terms of the series on either side of our differential equation we obtain

$$\begin{aligned} a_1 + (2)a_2 x + (3)a_3 x^2 + (4)a_4 x^3 + \cdots + (k+1)a_{k+1} x^k + \cdots \\ = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k + \cdots. \end{aligned}$$

Equating the constant terms gives us $a_1 = a_0$.

- c. Equating the degree 1 terms gives us $2a_2 = a_1$ or $a_2 = \frac{a_1}{2}$. Since $a_1 = a_0$, we have $a_2 = \frac{a_0}{2}$.
- d. Equating the degree 2 terms gives us $3a_3 = a_2$ or $a_3 = \frac{a_2}{3}$. Since $a_2 = \frac{a_0}{2}$, we have $a_3 = \frac{a_0}{3!}$.

e. Equating the degree 3 terms gives us $4a_4 = a_3$ or $a_4 = \frac{a_3}{4}$. Since $a_3 = \frac{a_0}{3!}$, we have $a_4 = \frac{a_0}{4!}$.

f. Equating the degree $k - 1$ terms gives us $ka_k = a_{k-1}$ or $a_k = \frac{a_{k-1}}{k}$. It appears that $a_{k-1} = \frac{a_0}{(k-1)!}$, so we have

$$a_k = \frac{a_0}{k!}.$$

g. Since $a_k = \frac{a_0}{k!}$ we have

$$y = a_0 \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

So the functions that satisfy the differential equation (8.6.1) are the exponential functions of the form $y = a_0 e^x$.

8.6.1 Power Series

As Example 8.6.1 shows, it can be useful to treat an unknown function as if it has a Taylor series, and then determine the coefficients from other information. In other words, we define a function as an infinite series of powers of x and then determine the coefficients based on something besides a formula for the function. This method allows us to approximate solutions to many different types of differential equations, even if we cannot solve them explicitly. This is different from our work with Taylor series since we are not using an original function f to generate the coefficients of the series.

Definition 8.6.2. A **power series** centered at $x = a$ is a function of the form

$$\sum_{k=0}^{\infty} c_k (x - a)^k \tag{8.6.2}$$

where $\{c_k\}$ is a sequence of real numbers and x is an independent variable.

A power series defines a function f whose domain is the set of x values for which the power series converges. We therefore write

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

It turns out that¹, on its interval of convergence, every power series is in fact the Taylor series of the function it defines, so all of the techniques we developed in the previous section can be applied to power series as well.

Example 8.6.3. Consider the power series defined by

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{2^k}.$$

¹See Exercise 8.6.4.4 in this section.

What are $f(1)$ and $f\left(\frac{3}{2}\right)$? Find a general formula for $f(x)$ and determine the values for which this power series converges.

Solution. If we evaluate f at $x = 1$ we obtain the series

$$\sum_{k=0}^{\infty} \frac{1}{2^k}$$

which is a geometric series with ratio $\frac{1}{2}$. So we can sum this series and find that

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

Similarly,

$$f\left(\frac{3}{2}\right) = \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k = \frac{1}{1 - \frac{3}{4}} = 4.$$

In general, $f(x)$ is a geometric series with ratio $\frac{x}{2}$, so

$$f(x) = \sum_{k=0}^{\infty} \left(\frac{x}{2}\right)^k = \frac{1}{1 - \frac{x}{2}} = \frac{2}{2 - x}$$

provided that $-1 < \frac{x}{2} < 1$ (which ensures that the ratio is less than 1 in absolute value). Thus, the power series that defines $f(x) = \frac{2}{2-x}$ converges for $-2 < x < 2$.

As we did for Taylor series, we define the interval of convergence of a power series (8.6.2) to be the set of values of x for which the series converges. And as we did with Taylor series, we typically use the Ratio Test to find the values of x for which the power series converges absolutely, and then check the endpoints separately if the radius of convergence is finite.

Example 8.6.4. Let $f(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^2}$. Determine the interval of convergence of this power series.

Solution. First we will plot some of the partial sums of this power series to get an idea of the interval of convergence. Let

$$S_n(x) = \sum_{k=1}^n \frac{x^k}{k^2}$$

for each $n \geq 1$. Figure 8.6.5 shows plots of $S_{10}(x)$ (in red), $S_{25}(x)$ (in blue), and $S_{50}(x)$ (in green).

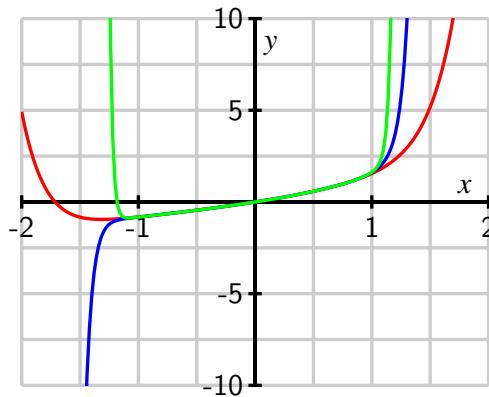


Figure 8.6.5: Graphs of some partial sums of the power series $\sum_{k=1}^{\infty} \frac{x^k}{k^2}$.

The behavior of S_{50} in particular suggests that $f(x)$ appears to be converging to a particular curve on the interval $(-1, 1)$, while growing without bound outside of that interval. Thus, the interval of convergence might be $-1 < x < 1$. To verify our conjecture, we apply the Ratio Test. Now,

$$a_k = \frac{x^k}{k^2},$$

so

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} &= \lim_{k \rightarrow \infty} \frac{\frac{|x|^{k+1}}{(k+1)^2}}{\frac{|x|^k}{k^2}} \\ &= \lim_{k \rightarrow \infty} |x| \left(\frac{k}{k+1} \right)^2 \\ &= |x| \lim_{k \rightarrow \infty} \left(\frac{k}{k+1} \right)^2 \\ &= |x|. \end{aligned}$$

Therefore, the Ratio Test tells us that $f(x)$ converges absolutely when $|x| < 1$ and diverges when $|x| > 1$. Because the Ratio Test is inconclusive when $|x| = 1$, we need to check $x = 1$ and $x = -1$ individually.

When $x = 1$, observe that

$$f(1) = \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

This is a p -series with $p > 1$, which we know converges. When $x = -1$, we have

$$f(-1) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}.$$

This is an alternating series, and since the sequence $\left\{\frac{1}{n^2}\right\}$ decreases to 0, the power series converges by the Alternating Series Test. Thus, the interval of convergence of this power series is $-1 \leq x \leq 1$.

Example 8.6.6. Determine the interval of convergence of each power series.

a. $\sum_{k=1}^{\infty} \frac{(x-1)^k}{3k}$

d. $\sum_{k=1}^{\infty} \frac{x^k}{(2k)!}$

b. $\sum_{k=1}^{\infty} kx^k$

c. $\sum_{k=1}^{\infty} \frac{k^2(x+1)^k}{4^k}$

e. $\sum_{k=1}^{\infty} k!x^k$

Hint.

- a. Small hints for each of the prompts above.

Answer.

a. $[0, 2)$.

b. $(-1, 1)$.

c. $(-5, 3)$.

d. $(-\infty, \infty)$.

e. $\{0\}$.

Solution.

a. We use the Ratio Test with $a_k = \frac{|x-1|^k}{3k}$:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{|x-1|^{k+1}}{3(k+1)}}{\frac{|x-1|^k}{3k}} &= \lim_{k \rightarrow \infty} \frac{3k|x-1|^{k+1}}{3(k+1)|x-1|^k} \\ &= |x-1| \lim_{k \rightarrow \infty} \frac{k}{k+1} \\ &= |x-1|. \end{aligned}$$

So the power series $\sum_{k=1}^{\infty} \frac{(x-1)^k}{3k}$ converges absolutely when $|x-1| < 1$ or when $0 < x < 2$ and diverges outside this interval. To completely determine the interval of convergence, we need to check what happens at the endpoints of this interval.

- When $x = 0$ our power series is $\sum_{k=1}^{\infty} \frac{(-1)^k}{3k}$ which is just a scalar multiple of the alternating harmonic series and so converges.
- When $x = 2$ our power series is $\sum_{k=1}^{\infty} \frac{1}{3k}$ which is just a scalar multiple of the harmonic series and so diverges.

Therefore, the interval of convergence of the power series $\sum_{k=1}^{\infty} \frac{(x-1)^k}{3k}$ is $[0, 2)$.

b. We use the Ratio Test with $a_k = k|x|^k$:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{(k+1)|x|^{k+1}}{k|x|^k} &= |x| \lim_{k \rightarrow \infty} \frac{k+1}{k} \\ &= |x|. \end{aligned}$$

So the power series $\sum_{k=1}^{\infty} kx^k$ converges absolutely when $|x| < 1$ or when $-1 < x < 1$ and diverges outside this interval. To completely determine the interval of convergence, we need to check what happens at the endpoints of this interval.

- When $x = -1$ our power series is $\sum_{k=1}^{\infty} (-1)^k k$. Since $k \rightarrow \infty$ as $k \rightarrow \text{inf}$, this series diverges by the Divergence Test.
- When $x = 1$ our power series is $\sum_{k=1}^{\infty} k$ which again diverges by the Divergence Test.

Therefore, the interval of convergence of the power series $\sum_{k=1}^{\infty} kx^k$ is $(-1, 1)$.

c. We use the Ratio Test with $a_k = \frac{k^2|x+1|^k}{4^k}$:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{(k+1)^2|x+1|^{k+1}}{4^{k+1}}}{\frac{k^2|x+1|^k}{4^k}} &= \lim_{k \rightarrow \infty} \frac{4^k (k+1)^2 |x+1|^{k+1}}{4^{k+1} k^2 |x+1|^k} \\ &= \frac{1}{4} |x+1| \lim_{k \rightarrow \infty} \left(\frac{k+1}{k} \right)^2 \\ &= \frac{1}{4} |x+1|. \end{aligned}$$

So the power series $\sum_{k=1}^{\infty} \frac{k^2(x+1)^k}{4^k}$ converges absolutely when $\frac{1}{4}|x+1| < 1$ or when $-5 < x < 3$ and diverges outside this interval. To completely determine the interval of convergence, we need to check what happens at the endpoints of this interval.

- When $x = -5$ our power series is $\sum_{k=1}^{\infty} (-1)^k k^2$. Since $k^2 \rightarrow \infty$ as $k \rightarrow \infty$, this series diverges by the Divergence Test.
- When $x = 3$ our power series is $\sum_{k=1}^{\infty} k^2$, which again diverges by the Divergence Test.

Therefore, the interval of convergence of the power series $\sum_{k=1}^{\infty} \frac{k^2(x+1)^k}{4^k}$ is $(-5, 3)$.

d. We use the Ratio Test with $a_k = \frac{|x|^k}{(2k)!}$:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\frac{|x|^{k+1}}{(2(k+1))!}}{\frac{|x|^k}{(2k)!}} &= \lim_{k \rightarrow \infty} |x| \frac{(2k)!}{(2(k+1))!} \\ &= |x| \lim_{k \rightarrow \infty} \frac{1}{(2k+2)(2k+1)} \\ &= 0. \end{aligned}$$

So the power series $\sum_{k=1}^{\infty} \frac{x^k}{(2k)!}$ converges absolutely on the interval $(-\infty, \infty)$.

- e. We use the Ratio Test with $a_k = k!|x|^k$:

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{(k+1)!|x|^{k+1}}{k!|x|^k} &= \lim_{k \rightarrow \infty} |x|(k+1) \\ &= \infty\end{aligned}$$

unless $x = 0$. So the interval of convergence of the power series $\sum_{k=1}^{\infty} \frac{x^k}{k!}$ is $\{0\}$.

8.6.2 Manipulating Power Series

We know power series expansions for important functions such as $\sin(x)$ and e^x . Often, we can use a known power series expansion to find a power series for a different, but related, function. The next example demonstrates one way to do this.

Example 8.6.7. Our goal in this example is to find a power series expansion for $f(x) = \frac{1}{1+x^2}$ centered at $x = 0$.

While we could use the methods of Section 8.5 and differentiate $f(x) = \frac{1}{1+x^2}$ several times to look for patterns and find the Taylor series for $f(x)$, we seek an alternate approach because of how complicated the derivatives of $f(x)$ quickly become.

- What is the Taylor series expansion for $g(x) = \frac{1}{1-x}$? What is the interval of convergence of this series?
- How is $g(-x^2)$ related to $f(x)$? Explain, and hence substitute $-x^2$ for x in the power series expansion for $g(x)$. Given the relationship between $g(-x^2)$ and $f(x)$, how is the resulting series related to $f(x)$?
- For which values of x will this power series expansion for $f(x)$ be valid? Why?

Hint.

- Small hints for each of the prompts above.

Answer.

- $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ for $-1 < x < 1$.
- $f(x) = g(-x^2) = \sum_{k=0}^{\infty} (-1)^k x^{2k}$.
- $-1 < x < 1$.

Solution.

- Recall that

$$g(x) = \frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

for $-1 < x < 1$.

- b. Substituting $-x^2$ for x in the power series expansion for $g(x)$ gives

$$\begin{aligned} f(x) &= g(-x^2) \\ &= \frac{1}{1 - (-x^2)} \\ &= \sum_{k=0}^{\infty} (-x^2)^k \\ &= \sum_{k=0}^{\infty} (-1)^k x^{2k}. \end{aligned}$$

- c. This power series expansion for $f(x)$ will be valid as long as $-1 < (-x)^2 < 1$ or for $-1 < x < 1$.

In a previous section we found several important Maclaurin series and their intervals of convergence. Here, we list these key functions and their corresponding expansions.

$$\begin{aligned} \sin(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} && \text{for } -\infty < x < \infty \\ \cos(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} && \text{for } -\infty < x < \infty \\ e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!} && \text{for } -\infty < x < \infty \\ \frac{1}{1-x} &= \sum_{k=0}^{\infty} x^k && \text{for } -1 < x < 1 \end{aligned}$$

As we saw in Example 8.6.7, we can use these known series to find other power series expansions for related functions such as $\sin(x^2)$, e^{5x^3} , and $\cos(x^5)$.

Example 8.6.8. Let f be the function given by the power series expansion

$$f(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}.$$

- a. Assume that we can differentiate a power series term by term, just like we can differentiate a (finite) polynomial. Use the fact that

$$f(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^k \frac{x^{2k}}{(2k)!} + \cdots$$

to find a power series expansion for $f'(x)$.

- b. Observe that $f(x)$ and $f'(x)$ have familiar Taylor series. What familiar functions are these? What known relationship does our work demonstrate?
- c. What is the series expansion for $f''(x)$? What familiar function is $f''(x)$?

Hint.

- a. Small hints for each of the prompts above.

Answer.

a. $f'(x) = -x + \frac{x^3}{3!} - \frac{x^5}{5!} + \cdots + (-1)^k \frac{x^{k-1}}{(k-1)!} + \cdots$

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

c. $f''(x) = -1 + \frac{x^2}{2!} - \frac{x^4}{4!} + \cdots$

Solution.

- a. Note that

$$f'(x) = -x + \frac{x^3}{3!} - \frac{x^5}{5!} + \cdots + (-1)^k \frac{x^{k-1}}{(k-1)!} + \cdots$$

- b. We recognize $f(x)$ as $\cos(x)$ and $f'(x)$ as $-\sin(x)$. This gives the known differentiation formula

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

- c. We see that

$$f''(x) = -1 + \frac{x^2}{2!} - \frac{x^4}{4!} + \cdots,$$

which is the series expansion for $-\cos(x)$ as expected

Our work in Example 8.6.7 holds more generally. The corresponding theorem, which we will not prove, states that we can differentiate a power series for a function f term by term and obtain the series expansion for f' , and similarly we can integrate a series expansion for a function f term by term and obtain the series expansion for $\int f(x) dx$. For both, the radius of convergence of the resulting series is the same as the original, though it is possible that the convergence status of the various series may differ at the endpoints. The formal statement of the Power Series Differentiation and Integration Theorem follows.

Power Series Differentiation and Integration Theorem

Suppose $f(x)$ has a power series expansion

$$f(x) = \sum_{k=0}^{\infty} c_k x^k$$

and that the series converges absolutely to $f(x)$ on the interval $-r < x < r$. Then, the

power series $\sum_{k=1}^{\infty} kc_k x^{k-1}$ obtained by differentiating the power series for $f(x)$ term by term converges absolutely to $f'(x)$ on the interval $-r < x < r$. That is,

$$f'(x) = \sum_{k=1}^{\infty} kc_k x^{k-1}, \text{ for } |x| < r.$$

Similarly, the power series $\sum_{k=0}^{\infty} c_k \frac{x^{k+1}}{k+1}$ obtained by integrating the power series for $f(x)$ term by term converges absolutely on the interval $-r < x < r$, and

$$\int f(x) dx = \sum_{k=0}^{\infty} c_k \frac{x^{k+1}}{k+1} + C, \text{ for } |x| < r.$$

This theorem validates the steps we took in Example 8.6.8. It tells us that we can differentiate and integrate term by term on the interior of the interval of convergence, but it does not tell us what happens at the endpoints of this interval. We always need to check what happens at the endpoints separately. More importantly, we can use the approach of differentiating or integrating a series term by term to find new series.

Example 8.6.9. Find a series expansion centered at $x = 0$ for $\arctan(x)$, as well as its interval of convergence.

Solution. While we could differentiate $\arctan(x)$ repeatedly and look for patterns in the derivative values at $x = 0$ in an attempt to find the Maclaurin series for $\arctan(x)$ from the definition, it turns out to be far easier to use a known series in an insightful way. In Example 8.6.7, we found that

$$\frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k}$$

for $-1 < x < 1$. Recall that

$$\frac{d}{dx} [\arctan(x)] = \frac{1}{1+x^2},$$

and therefore

$$\int \frac{1}{1+x^2} dx = \arctan(x) + C.$$

It follows that we can integrate the series for $\frac{1}{1+x^2}$ term by term to obtain the power series expansion for $\arctan(x)$. Doing so, we find that

$$\begin{aligned} \arctan(x) &= \int \left(\sum_{k=0}^{\infty} (-1)^k x^{2k} \right) dx \\ &= \sum_{k=0}^{\infty} \left(\int (-1)^k x^{2k} dx \right) \\ &= \left(\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} \right) + C. \end{aligned}$$

The Power Series Differentiation and Integration Theorem tells us that this equality is valid for at least $-1 < x < 1$.

To find the value of the constant C , we can use the fact that $\arctan(0) = 0$. So

$$0 = \arctan(0) = \left(\sum_{k=0}^{\infty} (-1)^k \frac{0^{2k+1}}{2k+1} \right) + C = C,$$

and we must have $C = 0$. Therefore,

$$\arctan(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} \quad (8.6.3)$$

for at least $-1 < x < 1$.

It is a straightforward exercise to check that the power series

$$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}$$

converges both when $x = -1$ and when $x = 1$; in each case, we have an alternating series with terms $\frac{1}{2k+1}$ that decrease to 0, and thus the interval of convergence for the series expansion for $\arctan(x)$ in Equation (8.6.3) is $-1 \leq x \leq 1$.

Example 8.6.10. Find a power series expansion for $\ln(1+x)$ centered at $x = 0$ and determine its interval of convergence.

Hint. Use the Taylor series expansion for $\frac{1}{1+x}$ centered at $x = 0$.

Answer. $\ln(1+x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1}$. for $-1 < x < 1$.

Solution. Using the Taylor series

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

for $-1 < x < 1$ we have that

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k.$$

Integrating the left and right sides of this last equation gives us

$$\ln(1+x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1} + C$$

for $-1 < x < 1$. Since $\ln(1) = 0$ we have that

$$0 = \sum_{k=0}^{\infty} (-1)^k \frac{0^{k+1}}{k+1} + C$$

and so $C = 0$. We conclude that

$$\ln(1+x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1}.$$

for $-1 < x < 1$.

8.6.3 Summary

- A power series is a series of the form

$$\sum_{k=0}^{\infty} a_k x^k.$$

- We can often assume a solution to a given problem can be written as a power series, then use the information in the problem to determine the coefficients in the power series. This method allows us to approximate solutions to certain problems using partial sums of the power series; that is, we can find approximate solutions that are polynomials.
- The connection between power series and Taylor series is that they are essentially the same thing: on its interval of convergence a power series is the Taylor series of its sum.

8.6.4 Exercises

1. (Finding coefficients in a power series expansion of a rational function) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
2. (Finding coefficients in a power series expansion of a function involving $\arctan(x)$) This exercise uses the WeBWorK Online Homework System and so is not available in the print copy of this book.
3. We can use power series to approximate definite integrals to which known techniques of integration do not apply. We will illustrate this in this exercise with the definite integral $\int_0^1 \sin(x^2) dx$.
 - a. Use the Taylor series for $\sin(x)$ to find the Taylor series for $\sin(x^2)$. What is the interval of convergence for the Taylor series for $\sin(x^2)$? Explain.
 - b. Integrate the Taylor series for $\sin(x^2)$ term by term to obtain a power series expansion for $\int \sin(x^2) dx$.
 - c. Use the result from part (b) to explain how to evaluate $\int_0^1 \sin(x^2) dx$. Determine the number of terms you will need to approximate $\int_0^1 \sin(x^2) dx$ to 3 decimal places.

Answer.

- a. $\sin(x^2) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2(k+1)}}{(2k+1)!}$, with interval of convergence $(-\infty, \infty)$.

b. $\int \sin(x^2) dx = \sum_{k=0}^{\infty} (-1)^k \frac{x^{4k+3}}{(2k+1)!(4k+3)} + C.$

c. $\int_0^1 \sin(x^2) dx = \sum_{k=0}^{\infty} (-1)^k \frac{1}{(2k+1)!(4k+3)}.$ Use $n = 1$ to generate the desired estimate.

Solution.

a. The Taylor series expansion for $\sin(x)$ centered at $x = 0$ is

$$\sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

Substituting x^2 for x gives us

$$\sin(x^2) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2(2k+1)}}{(2k+1)!},$$

as the Taylor series for $\sin(x^2)$ centered at $x = 0$. Since the interval of convergence for the Taylor series for $\sin(x)$ is $(-\infty, \infty)$, it follows that the interval of convergence of the Taylor series for $\sin(x^2)$ is also $(-\infty, \infty)$.

b. We can integrate a Taylor series term by term on its interval of convergence, so

$$\begin{aligned} \int \sin(x^2) dx &= \int \sum_{k=0}^{\infty} (-1)^k \frac{x^{2(2k+1)}}{(2k+1)!} dx \\ &= \sum_{k=0}^{\infty} \int (-1)^k \frac{x^{2(2k+1)}}{(2k+1)!} dx \\ &= \sum_{k=0}^{\infty} \int (-1)^k \frac{x^{4k+2}}{(2k+1)!} dx \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{4k+3}}{(2k+1)!(4k+3)} + C. \end{aligned}$$

c. To evaluate $\int_0^1 \sin(x^2) dx$, we evaluate the antiderivative $\sum_{k=0}^{\infty} (-1)^k \frac{x^{4k+3}}{(2k+1)!(4k+3)}$ at 1 and 0 and subtract. So

$$\begin{aligned} \int_0^1 \sin(x^2) dx &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{4k+3}}{(2k+1)!(4k+3)} \Big|_0^1 \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{1}{(2k+1)!(4k+3)}. \end{aligned}$$

The resulting series is an alternating series, so we can find a value of n that makes the n th partial sum of the series approximate the sum of the series to 3 decimal places by determining a value of n such that

$$a_{n+1} = \frac{1}{(2(n+1)+1)!(4(n+1)+3)} < 0.001.$$

This is not an inequality that we can solve algebraically, so we use trial and error. Doing so shows that $n = 1$ will work.

4. There is an important connection between power series and Taylor series. Suppose f is defined by a power series centered at 0 so that

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

- Determine the first 4 derivatives of f evaluated at 0 in terms of the coefficients a_k .
- Show that $f^{(n)}(0) = n!a_n$ for each positive integer n .
- Explain how the result of (b) tells us the following:

On its interval of convergence, a power series is the Taylor series of its sum.

Answer.

-
- Then

$$\begin{aligned} f'(x) &= \sum_{k=1}^{\infty} k a_k x^{k-1} \\ f''(x) &= \sum_{k=2}^{\infty} k(k-1) a_k x^{k-2} \\ f^{(3)}(x) &= \sum_{k=3}^{\infty} k(k-1)(k-2) a_k x^{k-3} \\ &\vdots \quad \vdots \\ f^{(n)}(x) &= \sum_{k=n}^{\infty} k(k-1)(k-2) \cdots (k-n+1) a_k x^{k-n} \\ &\vdots \quad \vdots \end{aligned}$$

So

$$\begin{aligned} f(0) &= a_0 \\ f'(0) &= a_1 \\ f''(0) &= 2!a_2 \\ f^{(3)}(0) &= 3!a_3 \\ &\vdots \quad \vdots \\ f^{(k)}(0) &= k!a_k \\ &\vdots \quad \vdots \end{aligned}$$

and

$$a_k = \frac{f^{(k)}(0)}{k!}$$

for each $k \geq 0$. But these are just the coefficients of the Taylor series expansion of f , which leads us to the following observation.

Solution.

a. Observe that

$$\begin{aligned}f'(x) &= \sum_{k=1}^{\infty} k a_k x^{k-1} \\f''(x) &= \sum_{k=2}^{\infty} k(k-1) a_k x^{k-2} \\f^{(3)}(x) &= \sum_{k=3}^{\infty} k(k-1)(k-2) a_k x^{k-3} \\f^{(4)}(x) &= \sum_{k=4}^{\infty} k(k-1)(k-2)(k-3) a_k x^{k-4}\end{aligned}$$

and therefore

$$\begin{aligned}f(0) &= a_0 \\f'(0) &= a_1 \\f''(0) &= 2!a_2 \\f^{(3)}(0) &= 3!a_3 \\f^{(4)}(0) &= 4!a_4.\end{aligned}$$

b. Since

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1)(k-2)\cdots(k-n+1) a_k x^{k-n}$$

every term of this series vanishes at $x = 0$ except the first. Thus it follows $f^{(n)}(0) = n(n-1)(n-2)\cdots(1)a_n$, so $f^{(n)}(0) = n!a_n$.

c. Since $a_k = \frac{f^{(k)}(0)}{k!}$ for each $k \geq 0$, we see that these are just the coefficients of the Taylor series expansion of f , and thus we get the unsurprising result that the coefficients of a power series are identical to the Taylor series of the power series.

5. In this exercise we will begin with a strange power series and then find its sum. The Fibonacci sequence $\{f_n\}$ is a famous sequence whose first few terms are

$$f_0 = 0, f_1 = 1, f_2 = 1, f_3 = 2, f_4 = 3, f_5 = 5, f_6 = 8, f_7 = 13, \dots,$$

where each term in the sequence after the first two is the sum of the preceding two terms. That is, $f_0 = 0$, $f_1 = 1$ and for $n \geq 2$ we have

$$f_n = f_{n-1} + f_{n-2}.$$

Now consider the power series

$$F(x) = \sum_{k=0}^{\infty} f_k x^k.$$

We will determine the sum of this power series in this exercise.

a. Explain why each of the following is true.

- $xF(x) = \sum_{k=1}^{\infty} f_{k-1}x^k$
- $x^2F(x) = \sum_{k=2}^{\infty} f_{k-2}x^k$

b. Show that

$$F(x) - xF(x) - x^2F(x) = x.$$

c. Now use the equation

$$F(x) - xF(x) - x^2F(x) = x$$

to find a simple form for $F(x)$ that doesn't involve a sum.

d. Use a computer algebra system or some other method to calculate the first 8 derivatives of $\frac{x}{1-x-x^2}$ evaluated at 0. Why shouldn't the results surprise you?

6. Airy's equation²

$$y'' - xy = 0, \quad (8.6.4)$$

can be used to model an undamped vibrating spring with spring constant x (note that y is an unknown function of x). So the solution to this differential equation will tell us the behavior of a spring-mass system as the spring ages (like an automobile shock absorber). Assume that a solution $y = f(x)$ has a Taylor series that can be written in the form

$$y = \sum_{k=0}^{\infty} a_k x^k,$$

where the coefficients are undetermined. Our job is to find the coefficients.

- Differentiate the series for y term by term to find the series for y' . Then repeat to find the series for y'' .
- Substitute your results from part (a) into the Airy equation and show that we can write Equation (8.6.4) in the form

$$\sum_{k=2}^{\infty} (k-1)ka_k x^{k-2} - \sum_{k=0}^{\infty} a_k x^{k+1} = 0. \quad (8.6.5)$$

- At this point, it would be convenient if we could combine the series on the left in (8.6.5), but one written with terms of the form x^{k-2} and the other with terms in the form x^{k+1} . Explain why

$$\sum_{k=2}^{\infty} (k-1)ka_k x^{k-2} = \sum_{k=0}^{\infty} (k+1)(k+2)a_{k+2} x^k. \quad (8.6.6)$$

- Now show that

$$\sum_{k=0}^{\infty} a_k x^{k+1} = \sum_{k=1}^{\infty} a_{k-1} x^k. \quad (8.6.7)$$

²The general differential equations of the form $y'' \pm k^2 xy = 0$ is called Airy's equation. These equations arise in many problems, such as the study of diffraction of light, diffraction of radio waves around an object, aerodynamics, and the buckling of a uniform column under its own weight.

(e) We can now substitute (8.6.6) and (8.6.7) into (8.6.5) to obtain

$$\sum_{n=0}^{\infty} (n+1)(n+2)a_{n+2}x^n - \sum_{n=1}^{\infty} a_{n-1}x^n = 0. \quad (8.6.8)$$

Combine the like powers of x in the two series to show that our solution must satisfy

$$2a_2 + \sum_{k=1}^{\infty} [(k+1)(k+2)a_{k+2} - a_{k-1}] x^k = 0. \quad (8.6.9)$$

(f) Use equation (8.6.9) to show the following:

- i. $a_{3k+2} = 0$ for every positive integer k ,
- ii. $a_{3k} = \frac{1}{(2)(3)(5)(6)\cdots(3k-1)(3k)}a_0$ for $k \geq 1$,
- iii. $a_{3k+1} = \frac{1}{(3)(4)(6)(7)\cdots(3k)(3k+1)}a_1$ for $k \geq 1$.

(g) Use the previous part to conclude that the general solution to the Airy equation (8.6.4) is

$$y = a_0 \left(1 + \sum_{k=1}^{\infty} \frac{x^{3k}}{(2)(3)(5)(6)\cdots(3k-1)(3k)} \right) + a_1 \left(x + \sum_{k=1}^{\infty} \frac{x^{3k+1}}{(3)(4)(6)(7)\cdots(3k)(3k+1)} \right).$$

Any values for a_0 and a_1 then determine a specific solution that we can approximate as closely as we like using this series solution.

Answer. The results from the various part of this exercise show that

$$y = a_0 \left(1 + \sum_{k=1}^{\infty} \frac{x^{3k}}{(2)(3)(5)(6)\cdots(3k-1)(3k)} \right) + a_1 \left(x + \sum_{k=1}^{\infty} \frac{x^{3k+1}}{(3)(4)(6)(7)\cdots(3k)(3k+1)} \right).$$

Solution.

a.

$$y' = \sum_{k=1}^{\infty} k a_k x^{k-1}$$

$$y'' = \sum_{k=2}^{\infty} (k-1) n a_k x^{k-2}.$$

b. We substitute our series from (a) into the Airy equation to obtain the equation

$$\sum_{k=2}^{\infty} (k-1) k a_k x^{k-2} + x \sum_{k=0}^{\infty} a_k x^k = 0.$$

Distributing the x in the second term on the left yields the desired equation.

c. Note that

$$\begin{aligned}
 & \sum_{k=2}^{\infty} (k-1)ka_k x^{k-2} \\
 &= (1)(2)a_2 + (2)(3)a_3 x + (3)(4)a_4 x^2 + (4)(5)a_5 x^3 + \cdots + (k-1)(k)a_{k+2} x^{k-2} + \cdots \\
 &= (0+1)(0+2)x^0 + (1+1)(1+2)x^1 + (2+1)(2+2)x^2 + \cdots + (k+1)(k+2)x^k + \cdots
 \end{aligned}$$

In other words, we can re-index this series by increasing every k by 2, or replacing $k-2$ with k , $k-1$ with $k+1$ and k with $k+2$. This gives us the next desired equation.

d. As we did in the previous part,

$$\begin{aligned}
 \sum_{k=0}^{\infty} a_k x^{k+1} &= a_0 x + a_1 x^2 + a_2 x^3 + \cdots + a_k x^{k+1} + \cdots \\
 &= a_0 x + a_1 x^2 + a_2 x^3 + \cdots + a_{k-1} x^k + \cdots \\
 &= \sum_{k=1}^{\infty} a_{k-1} x^k.
 \end{aligned}$$

e. Notice that we have like powers of x in the two series, so we can combine them and obtain

$$\begin{aligned}
 0 &= \sum_{k=0}^{\infty} (k+1)(k+2)a_{k+2} x^k + \sum_{k=1}^{\infty} a_{k-1} x^k \\
 &= \left[(1)(2)a_2 + \sum_{k=1}^{\infty} (k+1)(k+2)a_{k+2} x^k \right] + \sum_{k=1}^{\infty} a_{k-1} x^k \\
 &= 2a_2 + \sum_{k=1}^{\infty} [(k+1)(k+2)a_{k+2} + a_{k-1}] x^k.
 \end{aligned}$$

f. Equation (8.6.9) implies that $a_2 = 0$ and $(k+1)(k+2)a_{k+2} + a_{k-1} = 0$ for all $k \geq 1$.

Solving for a_{k+2} in the second equation shows that

$$\begin{aligned}
 a_2 &= 0 \\
 a_{k+2} &= -\frac{1}{(k+1)(k+2)} a_{k-1} \text{ for } k \geq 1.
 \end{aligned}$$

These last equations are called *recurrence relations* and allow us to write every coefficient of y in terms of a_0 and a_1 . For example, $k = 1$ shows that $a_3 = \frac{1}{(2)(3)}a_0 = \frac{1}{6}a_0$. We can continue in this way to obtain the first 10 coefficients in terms of a_0 and a_1 :

$$\begin{aligned}
 a_3 &= \frac{1}{(2)(3)}a_0 \\
 a_4 &= \frac{1}{(3)(4)}a_1 \\
 a_5 &= \frac{1}{(4)(5)}a_2 = 0
 \end{aligned}$$

$$\begin{aligned}
 a_6 &= \frac{1}{(5)(6)}a_3 = \frac{1}{(5)(6)}\left(\frac{1}{(2)(3)}a_0\right) = \frac{1}{(2)(3)(5)(6)}a_0 \\
 a_7 &= \frac{1}{(6)(7)}a_4 = \frac{1}{(6)(7)}\left(\frac{1}{(3)(4)}a_1\right) = \frac{1}{(3)(4)(6)(7)}a_1 \\
 a_8 &= \frac{1}{(7)(8)}a_5 = 0 \\
 a_9 &= \frac{1}{(8)(9)}a_6 = \frac{1}{(8)(9)}\left(\frac{1}{(2)(3)(5)(6)}a_0\right) = \frac{1}{(2)(3)(5)(6)(8)(9)}a_0 \\
 a_{10} &= \frac{1}{(9)(10)}a_7 = \frac{1}{(9)(10)}\left(\frac{1}{(3)(4)(6)(7)}a_1\right) = \frac{1}{(3)(4)(6)(7)(9)(10)}a_1.
 \end{aligned}$$

It may not be obvious, but there is a pattern.

- All of the terms involving a_2 are 0. These terms are a_2, a_5, a_8 , etc. The subscripts of these terms are all of the form $3k + 2$. So $a_{3k+2} = 0$ for every positive integer k .
- The terms that involve a_0 have the form a_3, a_6, a_9 , etc. and are all of the form a_{3k} for positive integers k . The pattern in the denominators of the coefficient for a_{3k} is $(2)(3)(5)(6) \cdots (3k-1)(3k)$. So

$$a_{3k} = \frac{1}{(2)(3)(5)(6) \cdots (3k-1)(3k)}a_0 \text{ for } k \geq 1.$$

- The terms that involve a_1 have the form a_4, a_7, a_{10} , etc. and are all of the form a_{3k+1} for positive integers k . The pattern in the denominators of the coefficient for a_{3k+1} is $(3)(4)(6)(7) \cdots (3k)(3k+1)$. So

$$a_{3k+1} = \frac{1}{(3)(4)(6)(7) \cdots (3k)(3k+1)}a_1 \text{ for } k \geq 1.$$

g. We can write the solution y in three pieces as

$$y = \sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} a_{3k} x^{3k} + \sum_{k=0}^{\infty} a_{3k+1} x^{3k+1} + \sum_{k=0}^{\infty} a_{3k+2} x^{3k+2}$$

and so our results from the previous part of this exercise show that

$$\begin{aligned}
 y &= a_0 \left(1 + \sum_{k=1}^{\infty} \frac{x^{3k}}{(2)(3)(5)(6) \cdots (3k-1)(3k)} \right) \\
 &\quad + a_1 \left(x + \sum_{k=1}^{\infty} \frac{x^{3k+1}}{(3)(4)(6)(7) \cdots (3k)(3k+1)} \right).
 \end{aligned}$$

Appendix A

A Short Table of Integrals

$$\text{a. } \int \frac{du}{a^2+u^2} = \frac{1}{a} \arctan \frac{u}{a} + C$$

$$\text{b. } \int \frac{du}{\sqrt{u^2 \pm a^2}} = \ln |u + \sqrt{u^2 \pm a^2}| + C$$

$$\text{c. } \int \sqrt{u^2 \pm a^2} du = \frac{u}{2} \sqrt{u^2 \pm a^2} \pm \frac{a^2}{2} \ln |u + \sqrt{u^2 \pm a^2}| + C$$

$$\text{d. } \int \frac{u^2 du}{\sqrt{u^2 \pm a^2}} = \frac{u}{2} \sqrt{u^2 \pm a^2} \mp \frac{a^2}{2} \ln |u + \sqrt{u^2 \pm a^2}| + C$$

$$\text{e. } \int \frac{du}{u \sqrt{u^2 + a^2}} = -\frac{1}{a} \ln \left| \frac{a + \sqrt{u^2 + a^2}}{u} \right| + C$$

$$\text{f. } \int \frac{du}{u \sqrt{u^2 - a^2}} = \frac{1}{a} \sec^{-1} \frac{u}{a} + C$$

$$\text{g. } \int \frac{du}{\sqrt{a^2 - u^2}} = \arcsin \frac{u}{a} + C$$

$$\text{h. } \int \sqrt{a^2 - u^2} du = \frac{u}{2} \sqrt{a^2 - u^2} + \frac{a^2}{2} \arcsin \frac{u}{a} + C$$

$$\text{i. } \int \frac{u^2}{\sqrt{a^2 - u^2}} du = -\frac{u}{2} \sqrt{a^2 - u^2} + \frac{a^2}{2} \arcsin \frac{u}{a} + C$$

$$\text{j. } \int \frac{du}{u \sqrt{a^2 - u^2}} = -\frac{1}{a} \ln \left| \frac{a + \sqrt{a^2 - u^2}}{u} \right| + C$$

$$\text{k. } \int \frac{du}{u^2 \sqrt{a^2 - u^2}} = -\frac{\sqrt{a^2 - u^2}}{a^2 u} + C$$

This appendix contains answers to all activities in the text. Answers for preview activities are not included. This appendix contains answers to all non- WeBWorK exercises in the text. For WeBWorK exercises, please use the [HTML](#) version of the text for access to answers and solutions.

Index

u-substitution, 532

absolute convergence, 867

acceleration, 113

alternating series, 861

alternating series estimation theorem, 865

alternating series test, 864

amplitude, xxviii

antiderivative, 392

general, 478, 493

graph, 475

antidifferentiation, 391

arc length, 608, 609

arcsine, 214

area, 600

under velocity function, 387

asymptote, 363

horizontal, 363

vertical, 363

autonomous, 702

average rate of change, 41

average value, 443

average value of a function, 442

average velocity, 27, 34

backward difference, 93

carrying capacity, 786

center of mass (continuous mass distribution), 649

center of mass (point-masses, 648

central difference, 93

chain rule, 197

codomain, 210

composition, 195

concave down, 112

concave up, 112

concavity, 111

conditional convergence, 867

constant multiple rule, 145

continuous, 1, 20

continuous at $x = c$, 21

converge

sequence, 803

convergence

absolute, 867

conditional, 867

convergent sequence, 803

cosecant, 185

cotangent, 185

critical number, 270

critical point, 270

critical value, 270

cusp, 130

decreasing, 103

definite integral

definition, 432, 433

density, 641

derivative

constant function, 143

cosine, 177

cotangent, 186

definition, 43, 67

exponential function, 143

inverse, 221

logarithm, 212

power function, 143

sine, 177

tangent, 186

difference quotient, 92

- differentiable, 44, 129
- differential equation, 505, 695
 - autonomous, 702
 - first order, 701
 - solution, 507
- disk method, 622
- distance traveled, 391
- diverge
 - sequence, 803
- Divergence Test, 838
- domain, 210
- dominates, 371
- equilibrium solution, 716
 - stable, 717
 - unstable, 717
- error, 581
- error functin, 516
- error function, 571
- Euler's Method, 730
 - error, 739
- evaluating the function, x
- extreme value, 268
- extreme value theorem, 292
- Fibonacci sequence, 797
- finite geometric series, 813
- first derivative test, 270
- foot-pound, 659
- forward difference, 93
- FTC, 460
- function, iii, 210
- function-derivative pair, 532
- Fundamental Theorem of Calculus
 - First, 511
 - Second, 513
- fundamental theorem of calculus, 458, 460
- geometric series, 814
 - common ratio, 815
- harmonic series, 839
- Hooke's Law, 661
- implicit function, 228
- improper integral, 678
 - converges, 681
 - diverges, 681
- unbounded integrand, 684
- unbounded region of integration, 678
- increasing, 103
- indefinite integral, 530
 - evaluate, 530
- indeterminate form, 12
- infinite series, 832
- infinity, 361
- inflection point, 277
- initial condition, 478
- input, iii
- instantaneous rate of change, 43, 92
- instantaneous velocity, 30
- integral function, 480
- integral sign, 433
- integral test, 839, 843
- integrand, 433
- integration by parts, 546
- interval of convergence, 894
- left limit, 15
- lemniscate, 228
- limit
 - one-sided, 15
- limit comparison test, 845
- limits of integration, 433
- local linearization, 251
- locally linear, 130
- logistic, 786
- logistic equation, 786
 - solution, 788
- Maclaurin series, 890
- mass, 641
- maximum
 - absolute, 267
 - global, 267
 - local, 268
 - relative, 268
- midline, xxviii
- midpoint rule
 - error, 581
- minimum
 - absolute, 267
 - global, 267
 - local, 268
 - relative, 268

- net signed area, 417
Newton's Law of Cooling, 702
Newton-meter, 659
- one-to-one, 210
onto, 210
output, iii
- parameters, xii
partial fractions, 564
partial sum, 817, 835
per capita growth rate, 784
period, xxvii
periodic function, xxvii
position, 26
power series, 910
 definition, 910
product rule, 157
- quotient rule, 159
- radian, xxix
ratio test, 847
related rates, 335
Riemann sum, 412
 left, 412
 middle, 413
 right, 413
right limit, 15
- secant, 185
second derivative, 109
second derivative test, 275
second fundamental theorem of calculus, 514
separable, 748, 749
sequence, 799
 term, 799
- sequence of partial sums, 835
series
 converges, 835
 diverges, 835
 geometric, 814
 sigma notation, 409
Simpson's rule, 586
slope, xi
slope field, 710, 712
slope-intercept form, xi
solid of revolution, 621
stable, 717
sum rule, 145
- tangent, 185
tangent line
 equation, 250
Taylor polynomial, 885
 error, 897
Taylor polynomials, 884
Taylor series, 889, 890
 interval of convergence, 894
 radius of convergence, 896
total change theorem, 463
trapezoid rule, 580
 error, 581
triangular numbers, 797
trigonometry, 184
 fundamental trigonometric identity, 184
two-parameter, xii
- unit circle, xxviii
unstable, 717
- washer method, 623
weighted average, 586, 645
work, 659, 660

This book was authored in PreTeXt.