lyryx with Open Texts

CALCULUS

Early Transcendentals

an Open Text

BASE TEXTBOOK

VERSION 2017– REVISION A

ADAPTABLE | ACCESSIBLE | AFFORDABLE

by Lyryx Learning
based on the original text by D. Guichard

Champions of Access to Knowledge

OPEN TEXT

All digital forms of access to our high-quality open texts are entirely FREE! All content is reviewed for excellence and is wholly adaptable; custom editions are produced by Lyryx for those adopting Lyryx assessment. Access to the original source files is also open to anyone!

ONLINE ASSESSMENT

We have been developing superior online formative assessment for more than 15 years. Our questions are continuously adapted with the content and reviewed for quality and sound pedagogy. To enhance learning, students receive immediate personalized feedback. Student grade reports and performance statistics are also provided.

SUPPORT

Access to our in-house support team is available 7 days/week to provide prompt resolution to both student and instructor inquiries. In addition, we work one-on-one with instructors to provide a comprehensive system, customized for their course. This can include adapting the text, managing multiple sections, and more!

INSTRUCTOR SUPPLEMENTS

Additional instructor resources are also freely accessible. Product dependent, these supplements include: full sets of adaptable slides and lecture notes, solutions manuals, and multiple choice question banks with an exam building tool.

Calculus – Early Transcendentals
an Open Text

BE A CHAMPION OF OER!

Contribute suggestions for improvements, new content, or errata:
A new topic
A new example
An interesting new question
A new or better proof to an existing theorem
Any other suggestions to improve the material

Contact Lyryx at info@lyryx.com with your ideas.

CONTRIBUTIONS

Jim Bailey, College of the Rockies
Mark Blenkinsop, Carleton University
Michael Cavers, University of Calgary
David Guichard, Whitman College
APEX Calculus: Gregory Hartman, Virginia Military Institute
Joseph Ling, University of Calgary

Lyryx Learning Team

Bruce Bauslaugh
Peter Chow
Nathan Friess
Stephanie Keyowski
Claude Laflamme
Martha Laflamme
Jennifer MacKenzie
Tamsyn Murnaghan
Bogdan Sava
Larissa Stone
Ryan Yee
Ehsun Zahedi

LICENSE

Calculus – Early Transcendentals

an Open Text

Base Text Revision History

Current Revision: Version 2017 — Revision A

Extensive edits, additions, and revisions have been completed by the editorial team at Lyryx Learning.
All new content (text and images) is released under the same license as noted above.

2017 A	• Lyryx: Front matter has been updated including cover page, copyright, and revision pages. • Lyryx: Examples 15.6, 16.11, 16.13, 16.3, 15.10, 16.20 and Exercises 15.4.3, 15.5.1, 16.6.1 have been rewritten. Several exercises from Vector Calculus have been removed. • Lyryx: Order and name of topics in Chapter 15 and Chapter 16 have been revised.
2016 B	• D. Guichard: New content developed for the Three Dimensions, Vector Functions, and Vector Calculus chapters. • Lyryx: Exercise numbering has been updated to restart with each section. • G. Hartman: New content on Riemann Sums is included, Section 6.1.1. This section was adapted by Lyryx from the section of the same name in APEX Calculus, Version 3.0, written by G. Hartman. T. Siemers and D. Chalishajar of the Virginia Military Instititue and B. Heinold of Mount Saint Mary's University also contributed to APEX Calculus. This material is released under Creative Commons license CC BY-NC (`https://creativecommons.org/licenses/by-nc/4.0/`). See `http://www.apexcalculus.com/` for more information and original version.
2016 A	• Lyryx: The layout and appearance of the text has been updated, including the title page and newly designed back cover.
2015 A	• J. Ling: Addition of new exercises and proofs throughout. • J. Ling: Revised arrangement of topics in the Application of Derivatives chapter. • J. Ling: Continuity section has been revised to include additional explanations of content and additional examples.
2014 A	• M. Cavers: Addition of new material and images particularly in the Review chapter. • M. Blenkinsop: Addition of content including Linear and Higher Order Approximations section.
2012 A	• **Original text** by D. Guichard of Whitman College, the single variable material is a modification and expansion of notes written and released by N. Koblitz of the University of Washington. That version also contains exercises and examples from *Elementary Calculus: An Approach Using Infinitesimals*, written by H. J. Keisler of the University of Wisconsin under a Creative Commons license (see `http://www.math.wisc.edu/~keisler/calc.html`). A. Schueller, B. Balof, and M. Wills all of Whitman College, have also contributed content. This material is released under the Creative Commons Attribution-NonCommercial-ShareAlike License (`http://creativecommons.org/licenses/by-nc-sa/3.0/`). See `http://communitycalculus.org` for more information.

Contents

Introduction

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

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

1. Carefully read each problem twice before writing anything.
2. Assign letters to quantities that are described only in words; draw a diagram if appropriate.
3. Decide which letters are constants and which are variables. A letter stands for a constant if its value remains the same throughout the problem.
4. Using mathematical notation, write down what you know and then write down what you want to find.
5. Decide what category of problem it is (this might be obvious if the problem comes at the end of a particular chapter, but will not necessarily be so obvious if it comes on an exam covering several chapters).
6. Double check each step as you go along; don't wait until the end to check your work.
7. Use common sense; if an answer is out of the range of practical possibilities, then check your work to see where you went wrong.

1. Review

Success in calculus depends on your background in algebra, trigonometry, analytic geometry and functions. In this chapter, we review many of the concepts you will need to know to succeed in this course.

1.1 Algebra

1.1.1. Sets and Number Systems

A **set** can be thought of as any collection of *distinct* objects considered as a whole. Typically, sets are represented using **set-builder notation** and are surrounded by braces. Recall that $(,)$ are called **parentheses** or **round brackets**; $[,]$ are called **square brackets**; and $\{,\}$ are called **braces** or **curly brackets**.

Example 1.1: Sets

The collection $\{a,b,1,2\}$ *is a set. It consists of the collection of four distinct objects, namely,* a, b, 1 *and* 2.

Let S be any set. We use the notation $\mathbf{x} \in \mathbf{S}$ to mean that x is an element *inside* of the set S, and the notation $\mathbf{x} \notin \mathbf{S}$ to mean that x is *not* an element of the set S.

Example 1.2: Set Membership

If $S = \{a,b,c\}$, *then* $a \in S$ *but* $d \notin S$.

The **intersection** between two sets S and T is denoted by $S \cap T$ and is the collection of all elements that belong to *both* S and T. The **union** between two sets S and T is denoted by $S \cup T$ and is the collection of all elements that belong to *either* S or T (or both).

Example 1.3: Union and Intersection

Let $S = \{a,b,c\}$ *and* $T = \{b,d\}$. *Then* $S \cap T = \{b\}$ *and* $S \cup T = \{a,b,c,d\}$. *Note that we do* not *write the element* b *twice in* $S \cup T$ *even though* b *is in both* S *and* T.

Numbers can be classified into sets called **number systems**.

$\mathbb{N}$	the **natural** numbers	$\{1,2,3,\ldots\}$
$\mathbb{Z}$	the **integers**	$\{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$
$\mathbb{Q}$	the **rational** numbers	Ratios of integers: $\left\{\frac{p}{q} : p,q \in \mathbb{Z}, q \neq 0\right\}$
$\mathbb{R}$	the **real** numbers	Can be written using a finite or infinite *decimal expansion*
$\mathbb{C}$	the **complex** numbers	These allow us to solve equations such as $x^2+1=0$

In the table, the set of rational numbers is written using set-builder notation. The colon, :, used in this manner means *such that*. Often times, a vertical bar | may also be used to mean *such that*. The expression $\left\{\frac{p}{q} : p,q \in \mathbb{Z}, q \neq 0\right\}$ can be read out loud as *the set of all fractions p over q such that p and q are both integers and q is not equal to zero*.

Example 1.4: Rational Numbers

The numbers $-\frac{3}{4}$, 2.647, 17, $0.\bar{7}$ *are all rational numbers. You can think of rational numbers as fractions of one integer over another. Note that* 2.647 *can be written as a fraction:*

$$2.647 = 2.647 \times \frac{1000}{1000} = \frac{2647}{1000}.$$

Also note that in the expression $0.\bar{7}$*, the bar over the* 7 *indicates that the* 7 *is repeated forever:*

$$0.77777777\ldots = \frac{7}{9}.$$

All rational numbers are real numbers with the property that their decimal expansion either *terminates* after a finite number of digits or begins to *repeat* the same finite sequence of digits over and over. Real numbers that are not rational are called **irrational**.

Example 1.5: Irrational Numbers

Some of the most common irrational numbers include:

- $\sqrt{2}$. *Can you prove this is irrational? (The proof uses a technique called contradiction.)*
- π. *Recall that* π *(***pi***) is defined as the ratio of the circumference of a circle to its diameter and can be approximated by* 3.14159265.
- e. *Sometimes called Euler's number,* e *can be approximated by* 2.718281828459. *We will review the definition of* e *in a later chapter.*

Let S and T be two sets. If every element of S is also an element of T, then we say S is a **subset** of T and write $S \subseteq T$. Furthermore, if S is a subset of T but not equal to T, we often write $S \subset T$. The five sets of numbers in the table give an increasing sequence of sets:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

That is, all natural numbers are also integers, all integers are also rational numbers, all rational numbers are also real numbers, and all real numbers are also complex numbers.

1.1.2. Law of Exponents

The Law of Exponents is a set of rules for simplifying expressions that governs the combination of exponents (powers). Recall that $\sqrt[n]{\ }$ denotes the nth root. For example $\sqrt[3]{8} = 2$ represents that the cube root of 8 is equal to 2.

Definition 1.6: Law of Exponents

Definitions

If m, n are positive integers, then:

1. $x^n = x \cdot x \cdot \ldots \cdot x$ (n times).
2. $x^0 = 1$, for $x \neq 0$.
3. $x^{-n} = \dfrac{1}{x^n}$, for $x \neq 0$.
4. $x^{m/n} = \sqrt[n]{x^m}$ or $\left(\sqrt[n]{x}\right)^m$, for $x \geq 0$.

Combining

1. $x^a x^b = x^{a+b}$.
2. $\dfrac{x^a}{x^b} = x^{a-b}$, for $x \neq 0$.
3. $(x^a)^b = x^{ab} = x^{ba} = \left(x^b\right)^a$.

Distributing

1. $(xy)^a = x^a y^a$, for $x \geq 0$, $y \geq 0$.
2. $\left(\dfrac{x}{y}\right)^a = \dfrac{x^a}{y^a}$, for $x \geq 0$, $y > 0$.

In the next example, the word *simplify* means *to make simpler* or to write the expression more compactly.

Example 1.7: Laws of Exponents

Simplify the following expression as much as possible assuming $x, y > 0$:

$$\frac{3x^{-2}y^3x}{y^2\sqrt{x}}.$$

Solution. Using the Law of Exponents, we have:

$$\begin{aligned}
\frac{3x^{-2}y^3x}{y^2\sqrt{x}} &= \frac{3x^{-2}y^3x}{y^2x^{\frac{1}{2}}}, && \text{since } \sqrt{x}=x^{\frac{1}{2}}, \\
&= \frac{3x^{-2}yx}{x^{\frac{1}{2}}}, && \text{since } \frac{y^3}{y^2}=y, \\
&= \frac{3y}{x^{\frac{3}{2}}}, && \text{since } \frac{x^{-2}x}{x^{\frac{1}{2}}}=\frac{x^{-1}}{x^{\frac{1}{2}}}=x^{-\frac{3}{2}}=\frac{1}{x^{\frac{3}{2}}}, \\
&= \frac{3y}{\sqrt{x^3}}, && \text{since } x^{\frac{3}{2}}=\sqrt{x^3}.
\end{aligned}$$

An answer of $3yx^{-3/2}$ is equally acceptable, and such an expression may prove to be computationally simpler, although a positive exponent may be preferred. ♣

1.1.3. The Quadratic Formula and Completing the Square

The technique of **completing the square** allows us to solve quadratic equations and also to determine the center of a circle/ellipse or the vertex of a parabola.

The main idea behind completing the square is to turn:

$$ax^2+bx+c$$

into

$$a(x-h)^2+k.$$

One way to complete the square is to use the following formula:

$$ax^2+bx+c=a\left(x+\frac{b}{2a}\right)^2-\frac{b^2}{4a^2}+c.$$

But this formula is a bit complicated, so some students prefer following the steps outlined in the next example.

Example 1.8: Completing the Square

Solve $2x^2+12x-32=0$ *by completing the square.*

Solution. In this instance, we will *not* divide by 2 first (usually you would) in order to demonstrate what you should do when the 'a' value is not 1.

$2x^2+12x-32=0$	Start with original equation.
$2x^2+12x=32$	Move the number over to the other side.
$2(x^2+6x)=32$	Factor out the a from the ax^2+bx expression.
$6 \rightarrow \frac{6}{2}=3 \rightarrow 3^2=9$	Take the number in front of x, **divide by** 2, then **square** it.
$2(x^2+6x+9)=32+2\cdot 9$	Add the result to both sides, taking $a=2$ into account.
$2(x+3)^2=50$	Factor the resulting perfect square trinomial.
	You have now completed the square!
$(x+3)^2=25 \rightarrow x=2 \text{ or } x=-8$	To solve for x, simply divide by $a=2$ and take square roots.

♣

Suppose we want to solve for x in the quadratic equation $ax^2+bx+c=0$, where $a\neq 0$. The solution(s) to this equation are given by the **quadratic formula**.

The Quadratic Formula

The solutions to $ax^2+bx+c=0$ (with $a\neq 0$) are $x=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$.

Proof. To prove the quadratic formula we use the technique of *completing the square*. The general technique involves taking an expression of the form x^2+rx and trying to find a number we can add so that we end up with a perfect square (that is, $(x+n)^2$). It turns out if you add $(r/2)^2$ then you can factor it as a perfect square.

For example, suppose we want to solve for x in the equation $ax^2+bx+c=0$, where $a\neq 0$. Then we can move c to the other side and divide by a (remember, $a\neq 0$ so we can divide by it) to get

$$x^2+\frac{b}{a}x=-\frac{c}{a}.$$

To write the left side as a perfect square we use what was mentioned previously. We have $r=(b/a)$ in this case, so we must add $(r/2)^2=(b/2a)^2$ to both sides

$$x^2+\frac{b}{a}x+\left(\frac{b}{2a}\right)^2=-\frac{c}{a}+\left(\frac{b}{2a}\right)^2.$$

We know that the left side can be factored as a perfect square

$$\left(x+\frac{b}{2a}\right)^2=-\frac{c}{a}+\left(\frac{b}{2a}\right)^2.$$

The right side simplifies by using the exponent rules and finding a common denominator

$$\left(x+\frac{b}{2a}\right)^2=\frac{-4ac+b^2}{4a^2}.$$

Taking the square root we get

$$x+\frac{b}{2a}=\pm\sqrt{\frac{-4ac+b^2}{4a^2}},$$

which can be rearranged as

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}.$$

In essence, the quadratic formula is just completing the square. ♣

1.1.4. Inequalities, Intervals and Solving Basic Inequalities

Inequality Notation

Recall that we use the symbols $<,>,\leq,\geq$ when writing an inequality. In particular,

- $a<b$ means a is to the *left* of b (that is, a is *strictly less* than b),
- $a\leq b$ means a is to the *left of or the same* as b (that is, a is *less than or equal* to b),
- $a>b$ means a is to the *right* of b (that is, a is *strictly greater* than b),
- $a\geq b$ means a is to the *right of or the same* as b (that is, a is *greater than or equal* to b).

To keep track of the difference between the symbols, some students use the following mnemonic.

Mnemonic

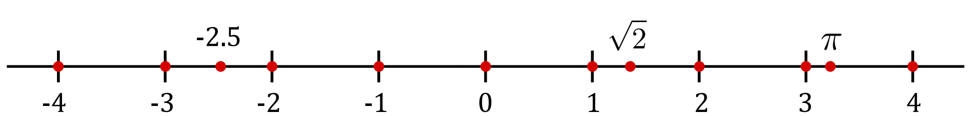
The $<$ symbol looks like a slanted L which stands for "*L*ess than".

Example 1.9: Inequalities

The following expressions are true:

$$1<2,\quad -5<-2,\quad 1\leq 2,\quad 1\leq 1,\quad 4\geq\pi>3,\quad 7.23\geq -7.23.$$

The real numbers are ordered and are often illustrated using the **real number line**:



Intervals

Assume a, b are real numbers with $a < b$ (i.e., a is strictly less than b). An **interval** is a set of every real number between two indicated numbers and may or may not contain the two numbers themselves. When describing intervals we use both round brackets and square brackets.

(1) Use of round brackets in intervals: (,). The notation **(a,b)** is what we call the **open interval from a to b** and consists of all the numbers between a and b, but does *not* include a or b. Using set-builder notation we write this as:

$$(a,b) = \{x \in \mathbb{R} : a < x < b\}.$$

We read $\{x \in \mathbb{R} : a < x < b\}$ as "the set of real numbers x such that x is greater than a and less than b" On the real number line we represent this with the following diagram:

Note that the circles on a and b are not shaded in, we call these **open circles** and use them to denote that a, b are *omitted* from the set.

(2) Use of square brackets in intervals: [,]. The notation **[a,b]** is what we call the **closed interval from a to b** and consists of all the numbers between a and b and *including* a and b. Using set-builder notation we write this as

$$[a,b] = \{x \in \mathbb{R} \,|\, a \leq x \leq b\}.$$

On the real number line we represent this with the following diagram:

Note that the circles on a and b are shaded in, we call these **closed circles** and use them to denote that a and b are *included* in the set.

To keep track of when to shade a circle in, you may find the following mnemonic useful:

Mnemonic

The round brackets (,) and non-shaded circle both form an "O" shape which stands for "Open and Omit".

Taking combinations of round and square brackets, we can write different possible types of intervals (we assume $a < b$):

$(a,b) = \{x \in \mathbb{R} : a < x < b\}$ 	$[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ 	$[a,b) = \{x \in \mathbb{R} : a \leq x < b\}$ 
$(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$ 	$(a,\infty) = \{x \in \mathbb{R} : x > a\}$ 	$[a,\infty) = \{x \in \mathbb{R} : x \geq a\}$ 
$(-\infty,b) = \{x \in \mathbb{R} : x < b\}$ 	$(-\infty,b] = \{x \in \mathbb{R} : x \leq b\}$ 	$(-\infty,\infty) = \mathbb{R} =$ all real numbers

Note: Any set which is bound at positive and/or negative infinity is an open interval.

Inequality Rules

Before solving inequalities, we start with the properties and rules of inequalities.

Inequality Rules

Add/subtract a number to both sides:

- If $a < b$, then $a + c < b + c$ and $a - c < b - c$.

Adding two inequalities of the same type:

- If $a < b$ and $c < d$, then $a + c < b + d$.
  Add the left sides together, add the right sides together.

Multiplying by a positive number:

- Let $c > 0$. If $a < b$, then $c \cdot a < c \cdot b$.

Multiplying by a negative number:

- Let $c < 0$. If $a < b$, then $c \cdot a > c \cdot b$.
  Note that we reversed the inequality symbol!

Similar rules hold for each of $\leq$, $>$ and $\geq$.

Solving Basic Inequalities

We can use the inequality rules to solve some simple inequalities.

Example 1.10: Basic Inequality

Find all values of x satisfying

$$3x + 1 > 2x - 3.$$

Write your answer in both interval and set-builder notation. Finally, draw a number line indicating your solution set.

Solution. Subtracting $2x$ from both sides gives $x + 1 > -3$. Subtracting 1 from both sides gives $x > -4$. Therefore, the solution is the interval $(-4, \infty)$. In set-builder notation the solution may be written as $\{x \in \mathbb{R} : x > -4\}$. We illustrate the solution on the number line as follows:

♣

Sometimes we need to split our inequality into two cases as the next example demonstrates.

Example 1.11: Double Inequalities

Solve the inequality

$$4 > 3x - 2 \geq 2x - 1.$$

Solution. We need both $4 > 3x-2$ and $3x-2 \geq 2x-1$ to be true:

$$\begin{array}{rcr} 4 > 3x-2 & \textit{and} & 3x-2 \geq 2x-1, \\ 6 > 3x & \textit{and} & x-2 \geq -1, \\ 2 > x & \textit{and} & x \geq 1, \\ x < 2 & \textit{and} & x \geq 1. \end{array}$$

Thus, we require $x \geq 1$ but also $x < 2$ to be true. This gives all the numbers between 1 and 2, including 1 but not including 2. That is, the solution to the inequality $4 > 3x-2 \geq 2x-1$ is the interval $[1,2)$. In set-builder notation this is the set $\{x \in \mathbb{R} : 1 \leq x < 2\}$. ♣

Example 1.12: Positive Inequality

Solve $4x - x^2 > 0$.

Solution. We provide two methods to solve this inequality.

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

Second method. Write $4x - x^2$ as $-(x^2-4x)$, and then complete the square, obtaining

$$-\left((x-2)^2-4\right) = 4-(x-2)^2.$$

For this to be positive we need $(x-2)^2 < 4$, which means that $x-2$ must be less than 2 and greater than -2: $-2 < x-2 < 2$. Adding 2 to everything gives $0 < x < 4$.

Both of these methods are equally correct; you may use either in a problem of this type. ♣

We next present another method to solve more complicated looking inequalities. In the next example we will solve a rational inequality by using a number line and test points. We follow the guidelines below.

Guidelines for Solving Rational Inequalities

1. Move everything to *one side* to get a 0 on the other side.
2. If needed, combine terms using a *common denominator*.
3. *Factor* the numerator and denominator.
4. Identify points where either the numerator or denominator is 0. Such points are called **split points**.
5. Draw a *number line* and indicate your split points on the number line. Draw *closed/open circles* for each split point depending on if that split point satisfies the inequality (division by zero is not allowed).
6. The split points will split the number line into subintervals. For each subinterval pick a *test point* and see if the expression in Step 3 is positive or negative. Indicate this with a $+$ or $-$ symbol on the number line for that subinterval.
7. Now write your answer in set-builder notation. Use the union symbol $\cup$ if you have multiple intervals in your solution.

Example 1.13: Rational Inequality

Write the solution to the following inequality using interval notation:

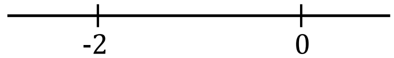
$$\frac{2-x}{2+x} \geq 1.$$

Solution. One method to solve this inequality is to multiply both sides by $2+x$, but because we do not know if $2+x$ is positive or negative we must split it into two cases (*Case 1:* $2+x>0$ and *Case 2:* $2+x<0$).

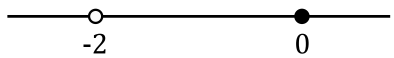
Instead we follow the guidelines for solving rational inequalities:

$$\begin{aligned}
\text{Start with original problem:} \quad & \frac{2-x}{2+x} \geq 1 \\
\text{Move everything to one side:} \quad & \frac{2-x}{2+x} - 1 \geq 0 \\
\text{Find a common denominator:} \quad & \frac{2-x}{2+x} - \frac{2+x}{2+x} \geq 0 \\
\text{Combine fractions:} \quad & \frac{(2-x)-(2+x)}{2+x} \geq 0 \\
\text{Expand numerator:} \quad & \frac{2-x-2-x}{2+x} \geq 0 \\
\text{Simplify numerator:} \quad & \frac{-2x}{2+x} \geq 0 \quad (*)
\end{aligned}$$

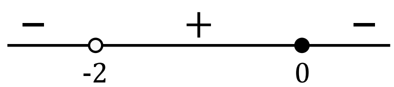
Now we have the numerator and denominator in fully factored form. The split points are $x = 0$ (makes the numerator 0) and $x = -2$ (makes the denominator 0). Let us draw a number line with the split points indicated on it:



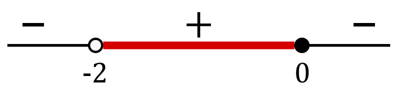
The point $x = 0$ is included since if we sub $x = 0$ into (*) we get $0 \geq 0$ which is true. The point $x = -2$ is not included since we cannot divide by zero. We indicate this with open/closed circles on the number line (remember that open means omit):



Now choosing a test point from each of the three subintervals we can determine if the expression $\frac{-2x}{2+x}$ is positive or negative. When $x = -3$, it is negative. When $x = -1$, it is positive. When $x = 1$, it is negative. Indicating this on the number line gives:



Since we wish to solve $\frac{-2x}{2+x} \geq 0$, we look at where the $+$ signs are and shade that area on the number line:



Since there is a closed circle at 0, we include it. Therefore, the solution is $(-2,0]$. ♣

Example 1.14: Rational Inequality

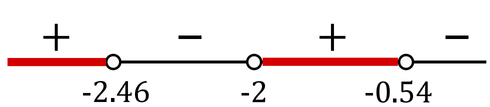
Write the solution to the following inequality using interval notation:

$$\frac{2}{x+2} > 3x+3.$$

Solution. We provide a brief outline of the solution. By subtracting $(3x+3)$ from both sides and using a common denominator of $x+2$, we can collect like terms and simplify to get:

$$\frac{-(3x^2+9x+4)}{x+2} > 0.$$

The denominator is zero when $x = -2$. Using the quadratic formula, the numerator is zero when $x = \frac{-9\pm\sqrt{33}}{6}$ (these two numbers are approximately -2.46 and -0.54). Since the inequality uses "$>$" and $0 > 0$ is false, we do not include any of the split points in our solution. After choosing suitable test points and determining the sign of $\frac{-(3x^2+9x+4)}{x+2}$ we have



Looking where the $+$ symbols are located gives the solution:

$$\left(-\infty, \frac{-9-\sqrt{33}}{6}\right) \bigcup \left(-2, \frac{-9+\sqrt{33}}{6}\right).$$

When writing the final answer we use *exact* expressions for numbers in mathematics, not approximations (unless stated otherwise). ♣

1.1.5. The Absolute Value

The **absolute value** of a number x is written as $|x|$ and represents the *distance* x is from zero. Mathematically, we define it as follows:

$$|x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

Thus, if x is a negative real number, then $-x$ is a positive real number. The absolute value does *not* just turn minuses into pluses. That is, $|2x-1| \neq 2x+1$. You should be familiar with the following properties.

Absolute Value Properties

1. $|x| \geq 0$.
2. $|xy| = |x||y|$.
3. $|1/x| = 1/|x|$ when $x \neq 0$.
4. $|-x| = |x|$.
5. $|x+y| \leq |x| + |y|$. This is called the **triangle inequality**.
6. $\sqrt{x^2} = |x|$.

Example 1.15: $\sqrt{x^2} = |x|$

Observe that $\sqrt{(-3)^2}$ *gives an answer of* 3*, not* -3*.*

When solving inequalities with absolute values, the following are helpful.

Case 1: $a > 0$**.**

- $|x| = a$ has solutions $x = \pm a$.
- $|x| \leq a$ means $x \geq -a$ **and** $x \leq a$ (that is, $-a \leq x \leq a$).
- $|x| < a$ means $x < -a$ **and** $x < a$ (that is, $-a < x < a$).
- $|x| \geq a$ means $x \leq -a$ **or** $x \geq a$.

- $|x| > a$ means $x < -a$ **or** $x > a$.

Case 2: $a < 0$.

- $|x| = a$ has no solutions.
- Both $|x| \leq a$ and $|x| < a$ have no solutions.
- Both $|x| \geq a$ and $|x| > a$ have solution set $\{x | x \in \mathbb{R}\}$.

Case 3: $a = 0$.

- $|x| = 0$ has solution $x = 0$.
- $|x| < 0$ has no solutions.
- $|x| \leq 0$ has solution $x = 0$.
- $|x| > 0$ has solution set $\{x \in \mathbb{R} | x \neq 0\}$.
- $|x| \geq 0$ has solution set $\{x | x \in \mathbb{R}\}$.

1.1.6. Solving Inequalities that Contain Absolute Values

We start by solving an equality that contains an absolute value. To do so, we recall that if $a \geq 0$ then the solution to $|x| = a$ is $x = \pm a$. In cases where we are not sure if the right side is positive or negative, we must perform a check at the end.

Example 1.16: Absolute Value Equality

Solve for x in $|2x+3| = 2-x$.

Solution. This means that either:

$$\begin{array}{ccc}
2x+3 = +(2-x) & or & 2x+3 = -(2-x) \\
2x+3 = 2-x & or & 2x+3 = -2+x \\
3x = -1 & or & x = -5 \\
x = -1/3 & or & x = -5
\end{array}$$

Since we do not know if the right side "$a = 2-x$" is positive or negative, we must perform a check of our answers and omit any that are incorrect.

If $x = -1/3$, then we have $LS = |2(-1/3)+3| = |-2/3+3| = |7/3| = 7/3$ and $RS = 2-(-1/3) = 7/3$. In this case $LS = RS$, so $x = -1/3$ is a solution.

If $x = -5$, then we have $LS = |2(-5)+3| = |-10+3| = |-7| = 7$ and $RS = 2-(-5) = 2+5 = 7$. In this case $LS = RS$, so $x = -5$ is a solution. ♣

We next look at absolute values and inequalities.

Example 1.17: Absolute Value Inequality

Solve $|x-5| < 7$.

Solution. This simply means $-7 < x-5 < 7$. Adding 5 to each gives $-2 < x < 12$. Therefore the solution is the interval $(-2, 12)$. ♣

In some questions you must be careful when multiplying by a negative number as in the next problem.

Example 1.18: Absolute Value Inequality

Solve $|2-z| < 7$.

Solution. This simply means $-7 < 2-z < 7$. Subtracting 2 gives: $-9 < -z < 5$. Now multiplying by -1 gives: $9 > z > -5$. *Remember to reverse the inequality signs!* We can rearrange this as $-5 < z < 9$. Therefore the solution is the interval $(-5, 9)$. ♣

Example 1.19: Absolute Value Inequality

Solve $|2-z| \geq 7$.

Solution. Recall that for $a > 0$, $|x| \geq a$ means $x \leq -a$ or $x \geq a$. Thus, either $2-z \leq -7$ *or* $2-z \geq 7$. Either $9 \leq z$ *or* $-5 \geq z$. Either $z \geq 9$ *or* $z \leq -5$. In interval notation, either z is in $[9, \infty)$ *or* z is in $(-\infty, -5]$. All together, we get our solution to be: $(-\infty, -5] \cup [9, \infty)$. ♣

In the previous two examples the *only* difference is that one had $<$ in the question and the other had $\geq$. Combining the two solutions gives the *entire* real number line!

Example 1.20: Absolute Value Inequality

Solve $0 < |x-5| \leq 7$.

Solution. We split this into two cases.

(1) For $0 < |x-5|$ note that we always have that an absolute value is positive or zero (i.e., $0 \leq |x-5|$ is always true). So, for this part, we need to avoid $0 = |x-5|$ from occurring. Thus, x *cannot* be 5, that is, $x \neq 5$.

(2) For $|x-5| \leq 7$, we have $-7 \leq x-5 \leq 7$. Adding 5 to each gives $-2 \leq x \leq 12$. Therefore the solution to $|x-5| \leq 7$ is the interval $[-2, 12]$.

To combine (1) and (2) we need combine $x \neq 5$ with $x \in [-2, 12]$. Omitting 5 from the interval $[-2, 12]$ gives our solution to be: $[-2, 5) \cup (5, 12]$. ♣

Exercises for 1.1

Exercise 1.1.1 *Simplify the following expressions as much as possible assuming $x, y > 0$:*

(a) $\dfrac{x^3 y^{-1/3}}{\sqrt[3]{y^2} x^2}$

(b) $\dfrac{3x^{-1/3} y^{-2} \sqrt[3]{x^4}}{\sqrt{9xy^{-3}}}$

(c) $\left(\dfrac{16x^2 y}{x^4}\right)^{1/2} \dfrac{\sqrt[3]{x^2}}{2\sqrt{y}}$

Exercise 1.1.2 *Find the constants a, b, c if the expression*

$$\frac{4x^{-1} y^2 \sqrt[3]{x}}{2x\sqrt{y}}$$

is written in the form $ax^b y^c$.

Exercise 1.1.3 *Find the roots of the quadratic equation*

$$x^2 - 2x - 24 = 0.$$

Exercise 1.1.4 *Solve the equation*

$$\frac{x}{4x-16} - 2 = \frac{1}{x-3}.$$

Exercise 1.1.5 *Solve the following inequalities. Write your answer as a union of intervals.*

(a) $3x + 1 > 6$

(b) $0 \leq 7x - 1 < 1$

(c) $\dfrac{x^2(x-1)}{(x+2)(x+3)^3} \leq 0$

(d) $x^2 + 1 > 0$

(e) $x^2 + 1 < 0$

(f) $x^2 + 1 > 2x$

(g) $x^3 > 4x$

(h) $x^3 \geq 4x^2$

(i) $\dfrac{1}{x} > 2$

(j) $\dfrac{x}{x+2} \leq \dfrac{2}{x-1}$

Exercise 1.1.6 *Solve the equation $|6x + 2| = 1$.*

Exercise 1.1.7 *Find solutions to the following absolute value inequalities. Write your answer as a union of intervals.*

(a) $|x| \geq 2$

(b) $|x-3| \leq 1$

(c) $|2x+5| \geq 4$

(d) $|x+2| < 3x-6$

(e) $|2x+5|+4 \geq 1$

(f) $5 < |x+1| < 8$

Exercise 1.1.8 *Solve the equation* $\sqrt{1-x}+x=1$.

1.2 Analytic Geometry

In what follows, we use the notation (x_1, y_1) to represent a point in the (x,y) coordinate system, also called the (x,y)-plane. Previously, we used (a,b) to represent an open interval. Notation often gets reused and abused in mathematics, but thankfully, it is usually clear from the context what we mean.

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

In applications, often letters other than x and y are used, and often different scales are chosen in the horizontal and vertical directions.

Example 1.21: Data Plot

Suppose you drop a coin from a window, and you want to study how its height above the ground changes from second to second. It is natural to let the letter t denote the time (the number of seconds since the object was released) and to let the letter h denote the height. For each t (say, at one-second intervals) you have a corresponding height h. This information can be tabulated, and then plotted on the (t,h) coordinate plane, as shown in figure 1.1.

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

Suppose we have two points A and B in the (x,y)-plane. We often want to know the change in x-coordinate (also called the "horizontal distance") in going from A to B. This is often written Δx, where the meaning of Δ (a capital delta in the Greek alphabet) is "change in". Thus, Δx can be read as "change in x" although it usually is read as "delta x". The point is that Δx denotes a single number, and should not be interpreted as "delta times x". Similarly, the "change in y" is written Δy and represents the difference between the y-coordinates of the two points. It is the vertical distance you have to move in going from A to B.

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

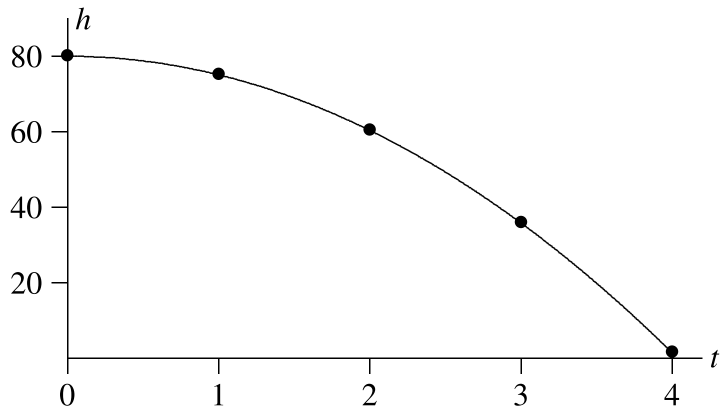


Figure 1.1: A data plot, height versus time.

Example 1.22: Change in x and y

If $A = (2,1)$ *and* $B = (3,3)$ *the change in* x *is*

$$\Delta x = 3 - 2 = 1$$

while the change in y *is*

$$\Delta y = 3 - 1 = 2.$$

The general formulas for the change in x and the change in y between a point (x_1, y_1) and a point (x_2, y_2) are:

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

Note that either or both of these might be negative.

1.2.1. Lines

If we have two *distinct* points $A(x_1, y_1)$ and $B(x_2, y_2)$, then we can draw one and only one straight line through both points. By the **slope** of this line we mean the ratio of Δy to Δx. The slope is often denoted by the letter m.

Slope Formula

The slope of the line joining the points (x_1, y_1) and (x_2, y_2) is:

$$m = \frac{\Delta y}{\Delta x} = \frac{(y_2 - y_1)}{(x_2 - x_1)} = \frac{\text{rise}}{\text{run}}.$$

Example 1.23: Slope of a Line Joining Two Points

The line joining the two points $(1,-2)$ *and* $(3,5)$ *has slope* $m = \dfrac{5-(-2)}{3-1} = \dfrac{7}{2}$.

The most familiar form of the equation of a straight line is:

$$y = mx + b.$$

Here m is the slope of the line: if you increase x by 1, the equation tells you that you have to increase y by m; and if you increase x by Δx, then y increases by $\Delta y = m\Delta x$. The number b is called the **y-intercept**, because it is where the line crosses the y-axis (when $x = 0$). If you know two points on a line, the formula $m = (y_2 - y_1)/(x_2 - x_1)$ gives you the slope. Once you know a point and the slope, then the y-intercept can be found by substituting the coordinates of either point in the equation: $y_1 = mx_1 + b$, i.e., $b = y_1 - mx_1$. Alternatively, one can use the **"point-slope" form** of the equation of a straight line: start with $(y - y_1)/(x - x_1) = m$ and then multiply to get

$$(y - y_1) = m(x - x_1),$$

the point-slope form. Of course, this may be further manipulated to get $y = mx - mx_1 + y_1$, which is essentially the **"$y = mx + b$" form**.

It is possible to find the equation of a line between two points directly from the relation $m = (y - y_1)/(x - x_1) = (y_2 - y_1)/(x_2 - x_1)$, which says "the slope measured between the point (x_1, y_1) and the point (x_2, y_2) is the same as the slope measured between the point (x_1, y_1) and any other point (x, y) on the line." For example, if we want to find the equation of the line joining our earlier points $A(2,1)$ and $B(3,3)$, we can use this formula:

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

Of course, this is really just the point-slope formula, except that we are not computing m in a separate step. We summarize the three common forms of writing a straight line below:

Slope-Intercept Form of a Straight Line

An equation of a line with slope m and y-intercept b is:

$$y = mx + b.$$

Point-Slope Form of a Straight Line

An equation of a line passing through (x_1, y_1) and having slope m is:

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

General Form of a Straight Line

Any line can be written in the form

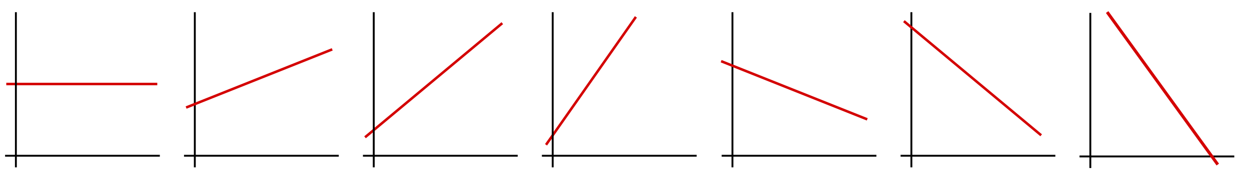
$$Ax + By + C = 0,$$

where A, B, C are constants and A, B are not both 0.

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

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

All of these possibilities are illustrated below.



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

It is often useful to find the x-intercept of a line $y = mx + b$. This is the x-value when $y = 0$. Setting $mx + b$ equal to 0 and solving for x gives: $x = -b/m$.

Example 1.24: Finding x-intercepts

To find x-intercept(s) of the line $y = 2x - 3$ we set $y = 0$ and solve for x:

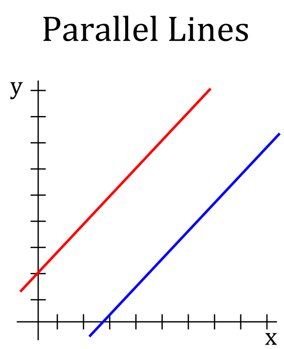
$$0 = 2x - 3 \qquad \rightarrow \qquad x = \frac{3}{2}.$$

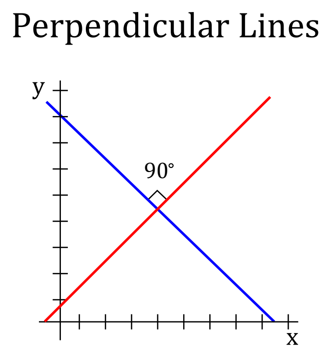
Thus, the line has an x-intercept of $3/2$.

It is often necessary to know if two lines are parallel or perpendicular. Let m_1 and m_2 be the slopes of the nonvertical lines L_1 and L_2. Then:

- L_1 and L_2 are **parallel** if and only if $m_1 = m_2$.
- L_1 and L_2 are **perpendicular** if and only if $m_2 = \dfrac{-1}{m_1}$. (Equivalently, $m_1 = \frac{-1}{m_2}$).

In the case of perpendicular lines, we say their slopes are *negative reciprocals*. Below is a visual representation of a pair of parallel lines and a pair of perpendicular lines.





Example 1.25: Equation of a Line

For each part below, find an equation of a line satisfying the requirements:

(a) *Through the two points* $(0,3)$ *and* $(-2,4)$.

(b) *With slope* 7 *and through point* $(1,-2)$.

(c) *With slope* 2 *and* y*-intercept* 4.

(d) *With* x*-intercept* 8 *and* y*-intercept* -3.

(e) *Through point* $(5,3)$ *and parallel to the line* $2x+4y+2=0$.

(f) *With* y*-intercept* 4 *and perpendicular to the line* $y=-\frac{2}{3}x+3$.

Solution. (a) We use the *slope formula* on $(x_1,y_1)=(0,3)$ and $(x_2,y_2)=(-2,4)$ to find m:

$$m=\frac{(4)-(3)}{(-2)-(0)}=\frac{1}{-2}=-\frac{1}{2}.$$

Now using the *point-slope formula* we get an equation to be:

$$y-3=-\frac{1}{2}(x-0) \quad \rightarrow \quad y=-\frac{1}{2}x+3.$$

(b) Using the *point-slope formula* with $m=7$ and $(x_1,y_1)=(1,-2)$ gives:

$$y-(-2)=7(x-1) \quad \rightarrow \quad y=7x-9.$$

(c) Using the *slope-intercept formula* with $m=2$ and $b=4$ we get $y=2x+4$.

(d) Note that the intercepts give us two points: $(x_1,y_1)=(8,0)$ and $(x_2,y_2)=(0,-3)$. Now follow the steps in part (a):

$$m=\frac{-3-0}{0-8}=\frac{3}{8}.$$

Using the *point-slope formula* we get an equation to be:

$$y-(-3)=\frac{3}{8}(x-0) \quad \rightarrow \quad y=\frac{3}{8}x-3.$$

(e) The line $2x+4y+2=0$ can be written as:

$$4y=-2x-2 \quad \rightarrow \quad y=-\frac{1}{2}x-\frac{1}{2}.$$

This line has slope $-1/2$. Since our line is *parallel* to it, we have $m=-1/2$. Now we have a point $(x_1,y_1)=(5,3)$ and slope $m=-1/2$, thus, the *point-slope formula* gives:

$$y-3=-\frac{1}{2}(x-5).$$

(f) The line $y=-\frac{2}{3}x+3$ has slope $m=-2/3$. Since our line is perpendicular to it, the slope of our line is the *negative reciprocal*, hence, $m=3/2$. Now we have $b=4$ and $m=3/2$, thus by the *slope-intercept formula*, an equation of the line is

$$y=\frac{3}{2}x+4.$$

♣

Example 1.26: Parallel and Perpendicular Lines

Are the two lines $7x+2y+3=0$ and $6x-4y+2=0$ perpendicular? Are they parallel? If they are not parallel, what is their point of intersection?

Solution. The first line is:

$$7x+2y+3=0 \quad \rightarrow \quad 2y=-7x-3 \quad \rightarrow \quad y=-\frac{7}{2}x-\frac{3}{2}.$$

It has slope $m_1=-7/2$. The second line is:

$$6x-4y+2=0 \quad \rightarrow \quad -4y=-6x-2 \quad \rightarrow \quad y=\frac{3}{2}x+\frac{1}{2}.$$

It has slope $m_2=3/2$. Since $m_1\cdot m_2\neq -1$ (they are not negative reciprocals), the lines are not perpendicular. Since $m_1\neq m_2$ the lines are not parallel.

We find points of intersection by setting y-values to be the same and solving. In particular, we have

$$-\frac{7}{2}x-\frac{3}{2}=\frac{3}{2}x+\frac{1}{2}.$$

Solving for x gives $x=-2/5$. Then substituting this into either equation gives $y=-1/10$. Therefore, the lines intersect at the point $(-2/5,-1/10)$. ♣

1.2.2. Distance between Two Points and Midpoints

Given two points (x_1,y_1) and (x_2,y_2), recall that their horizontal distance from one another is $\Delta x = x_2 - x_1$ and their vertical distance from one another is $\Delta y = y_2 - y_1$. Actually, the word "distance" normally denotes "positive distance". Δx and Δy are *signed* distances, but this is clear from context. The (positive) distance from one point to the other is the length of the hypotenuse of a right triangle with legs $|\Delta x|$ and $|\Delta y|$, as shown in figure 1.2. The Pythagorean Theorem states that the distance between the two points is the square root of the sum of the squares of the horizontal and vertical sides:

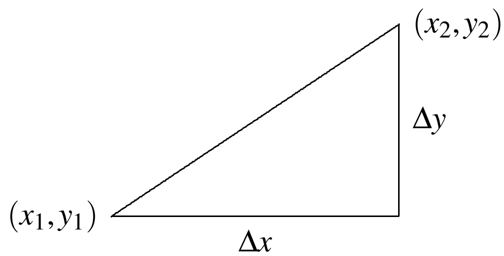


Figure 1.2: Distance between two points (here, Δx and Δy are positive).

Distance Formula

The distance between points (x_1,y_1) and (x_2,y_2) is

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

Example 1.27: Distance Between Two Points

The distance, d, between points $A(2,1)$ and $B(3,3)$ is

$$d = \sqrt{(3-2)^2 + (3-1)^2} = \sqrt{5}.$$

As a special case of the distance formula, suppose we want to know the distance of a point (x,y) to the origin. According to the distance formula, this is

$$\sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2+y^2}.$$

A point (x,y) is at a distance r from the origin if and only if $\sqrt{x^2+y^2} = r$, or, if we square both sides: $x^2+y^2 = r^2$. As we will see, this is the equation of the circle of radius, r, centered at the origin.

Furthermore, given two points we can determine the **midpoint** of the line segment joining the two points.

Midpoint Formula

The midpoint of the line segment joining two points (x_1,y_1) and (x_2,y_2) is the point with coordinates:

$$\text{midpoint} = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right).$$

Example 1.28: Midpoint of a Line Segment

Find the midpoint of the line segment joining the given points: $(1,0)$ *and* $(5,-2)$.

Solution. Using the *midpoint formula* on $(x_1,y_1)=(1,0)$ and $(x_2,y_2)=(5,-2)$ we get:

$$\left(\frac{(1)+(5)}{2}, \frac{(0)+(-2)}{2}\right) = (3,-1).$$

Thus, the midpoint of the line segment occurs at $(3,-1)$. ♣

1.2.3. Conics

In this section we review equations of parabolas, circles, ellipses and hyperbolas. We will give the equations of various conics in **standard form** along with a sketch. A useful mnemonic is the following.

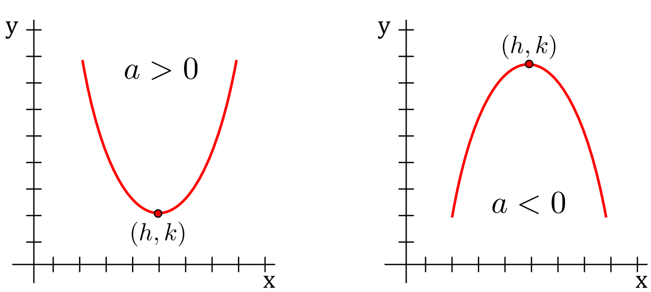
Mnemonic

In each conic formula presented, the terms '$x-h$' and '$y-k$' will always appear. The point (h,k) will alway represent either the centre or vertex of the particular conic.

Note that h or k (or both) may equal 0.

Vertical Parabola: The equation of a vertical parabola is:

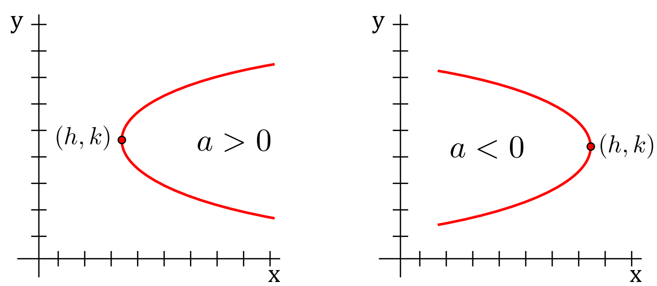
$$y-k=a(x-h)^2$$



- (h,k) is the *vertex* of the parabola.
- a is the vertical *stretch factor*.
- If $a > 0$, the parabola opens *upward*.
- If $a < 0$, the parabola opens *downward*.

Horizontal Parabola: The equation of a horizontal parabola is:

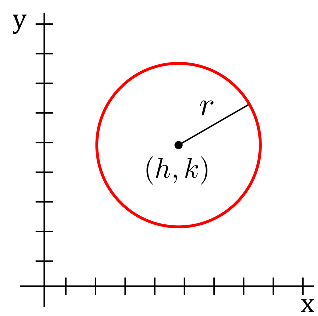
$$x - h = a(y-k)^2$$



- (h,k) is the *vertex* of the parabola.
- a is the horizontal *stretch factor*.
- If $a > 0$, the parabola opens *right*.
- If $a < 0$, the parabola opens *left*.

Circle: The equation of a circle is:

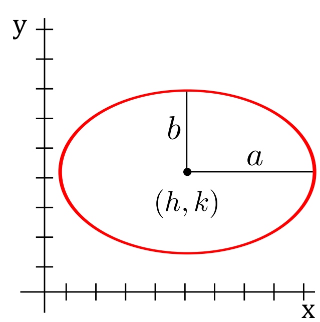
$$(x-h)^2 + (y-k)^2 = r^2$$



- (h,k) is the *centre* of the circle.
- r is the *radius* of the circle.

Ellipse: The equation of an ellipse is:

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



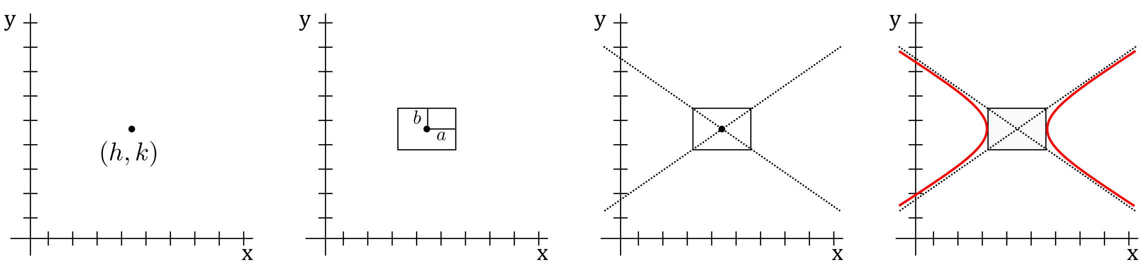
- (h,k) is the *centre* of the ellipse.
- a is the *horizontal distance* from the centre to the edge of the ellipse.
- b is the *vertical distance* from the centre to the edge of the ellipse.

Horizontal Hyperbola: The equation of a horizontal hyperbola is:

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

- (h,k) is the *centre* of the hyperbola.
- a,b are the *reference box* values. The box has a centre of (h,k).
- a is the *horizontal distance* from the centre to the edge of the box.
- b is the *vertical distance* from the centre to the edge of the box.

Given the equation of a horizontal hyperbola, one may sketch it by first placing a dot at the point (h,k). Then draw a box around (h,k) with horizontal distance a and vertical distance b to the edge of the box. Then draw dotted lines (called the **asymptotes** of the hyperbola) through the corners of the box. Finally, sketch the hyperbola in a horizontal direction as illustrated below.

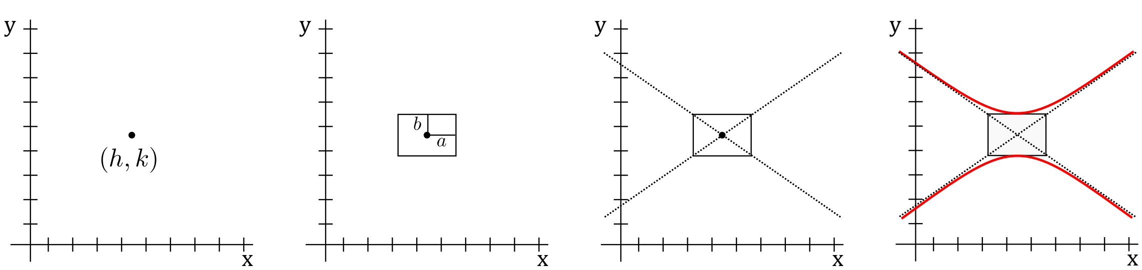


Vertical Hyperbola: The equation of a vertical hyperbola is:

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

- (h,k) is the *centre* of the hyperbola.
- a,b are the *reference box* values. The box has a centre of (h,k).
- a is the *horizontal distance* from the centre to the edge of the box.
- b is the *vertical distance* from the centre to the edge of the box.

Given the equation of a vertical hyperbola, one may sketch it by following the same steps as with a horizontal hyperbola, but sketching the hyperbola going in a vertical direction.



Determining the Type of Conic

An equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

gives rise to a graph that can be generated by performing a conic section (parabolas, circles, ellipses, hyperbolas). Note that the Bxy term involves conic rotation. The Dx, Ex, and F terms affect the vertex and centre. For simplicity, we will omit the Bxy term. To determine the type of graph we focus our analysis on the values of A and C.

- If $A = C$, the graph is a *circle*.
- If $AC > 0$ (and $A \neq C$), the graph is an *ellipse*.
- If $AC = 0$, the graph is a *parabola*.
- If $AC < 0$, the graph is a *hyperbola*.

Example 1.29: Center and Radius of a Circle

Find the centre and radius of the circle $y^2 + x^2 - 12x + 8y + 43 = 0$.

Solution. We need to complete the square twice, once for the x terms and once for the y terms. We'll do both at the same time. First let's collect the terms with x together, the terms with y together, and move the number to the other side.

$$(x^2 - 12x) + (y^2 + 8y) = -43$$

We add 36 to both sides for the x term ($-12 \to \frac{-12}{2} = -6 \to (-6)^2 = 36$), and 16 to both sides for the y term ($8 \to \frac{8}{2} = 4 \to (4)^2 = 16$):

$$(x^2 - 12x + 36) + (y^2 + 8y + 16) = -43 + 36 + 16$$

Factoring gives:

$$(x-6)^2 + (y+4)^2 = 3^2.$$

Therefore, the centre of the circle is $(6, -4)$ and the radius is 3. ♣

Example 1.30: Type of Conic

What type of conic is $4x^2 - y^2 - 8x + 8 = 0$? *Put it in standard form.*

Solution. Here we have $A = 4$ and $C = -1$. Since $AC < 0$, the conic is a hyperbola. Let us complete the square for the x and y terms. First let's collect the terms with x together, the terms with y together, and move the number to the other side.

$$(4x^2 - 8x) - y^2 = -8$$

Now we factor out 4 from the x terms.

$$4(x^2 - 2x) - y^2 = -8$$

Notice that we don't need to complete the square for the y terms (it is already completed!). To complete the square for the x terms we add 1 since $-2 \to \frac{-2}{2} = -1 \to (-1)^2 = 1$, taking into consideration that the a value is 4:

$$4(x^2 - 2x + 1) - y^2 = -8 + 4 \cdot 1$$

Factoring gives:

$$4(x-1)^2 - y^2 = -4$$

A hyperbola in standard form has ± 1 on the right side and a positive x^2 on the left side, thus, we must divide by 4:

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

Now we can see that the equation represents a vertical hyperbola with centre $(1,0)$ (and with a value $\sqrt{1} = 1$, and b value $\sqrt{4} = 2$). ♣

Example 1.31: Equation of Parabola

Find an equation of the parabola with vertex $(1,-1)$ that passes through the points $(-4,24)$ and $(7,35)$.

Solution. We first need to determine if it is a vertical parabola or horizontal parabola. See figure 1.3 for a sketch of the three points $(1,-1)$, $(-4,24)$ and $(7,35)$ in the xy-plane. Note that the vertex is $(1,-1)$.

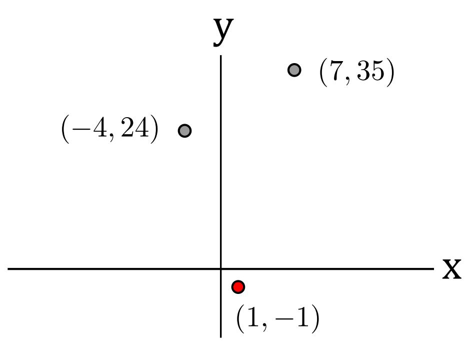


Figure 1.3: Figure for Example 1.31

Given the location of the vertex, the parabola cannot open downwards. It also cannot open left or right (because the vertex is between the other two points - if it were to open to the right, every other point would need to be to the right of the vertex; if it were to open to the left, every other point would need to be to the left of the vertex). Therefore, the parabola must open upwards and it is a vertical parabola. It has an equation of

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

As the vertex is $(h,k) = (1,-1)$ we have:

$$y - (-1) = a(x-1)^2$$

To determine a, we substitute one of the points into the equation and solve. Let us substitute the point $(x,y) = (-4,24)$ into the equation:

$$24 - (-1) = a(-4-1)^2 \quad \to \quad 25 = 25a \quad \to \quad a = 1.$$

Therefore, the equation of the parabola is:

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

Note that if we substituted $(7,35)$ into the equation instead, we would also get $a=1$. ♣

Exercises for 1.2

Exercise 1.2.1 *Find the equation of the line in the form $y=mx+b$:*

(a) *through* $(1,1)$ *and* $(-5,-3)$

(b) *through* $(-1,2)$ *with slope* -2

(c) *through* $(-1,1)$ *and* $(5,-3)$

(d) *through* $(2,5)$ *and parallel to the line* $3x+9y+6=0$

(e) *with x-intercept 5 and perpendicular to the line* $y=2x+4$

Exercise 1.2.2 *Change the following equations to the form $y=mx+b$, graph the line, and find the y-intercept and x-intercept.*

(a) $y-2x=2$

(b) $x+y=6$

(c) $x=2y-1$

(d) $3=2y$

(e) $2x+3y+6=0$

Exercise 1.2.3 *Determine whether the lines $3x+6y=7$ and $2x+4y=5$ are parallel.*

Exercise 1.2.4 *Suppose a triangle in the (x,y)-plane has vertices $(-1,0)$, $(1,0)$ and $(0,2)$. Find the equations of the three lines that lie along the sides of the triangle in $y=mx+b$ form.*

Exercise 1.2.5 *Let x stand for temperature in degrees Celsius (centigrade), and let y stand for temperature in degrees Fahrenheit. A temperature of $0°C$ corresponds to $32°F$, and a temperature of $100°C$ corresponds to $212°F$. Find the equation of the line that relates temperature Fahrenheit y to temperature Celsius x in the form $y=mx+b$. Graph the line, and find the point at which this line intersects $y=x$. What is the practical meaning of this point?*

Exercise 1.2.6 *A car rental firm has the following charges for a certain type of car: \$25 per day with 100 free miles included, \$0.15 per mile for more than 100 miles. Suppose you want to rent a car for one day, and you know you'll use it for more than 100 miles. What is the equation relating the cost y to the number of miles x that you drive the car?*

Exercise 1.2.7 *A photocopy store advertises the following prices: 5c per copy for the first 20 copies, 4c per copy for the 21st through 100th copy, and 3c per copy after the 100th copy. Let x be the number of copies, and let y be the total cost of photocopying. (a) Graph the cost as x goes from 0 to 200 copies. (b) Find the equation in the form* $y = mx + b$ *that tells you the cost of making x copies when x is more than 100.*

Exercise 1.2.8 *Market research tells you that if you set the price of an item at \$1.50, you will be able to sell 5000 items; and for every 10 cents you lower the price below \$1.50 you will be able to sell another 1000 items. Let x be the number of items you can sell, and let P be the price of an item. (a) Express P linearly in terms of x, in other words, express P in the form* $P = mx + b$*. (b) Express x linearly in terms of P.*

Exercise 1.2.9 *An instructor gives a 100-point final exam, and decides that a score 90 or above will be a grade of 4.0, a score of 40 or below will be a grade of 0.0, and between 40 and 90 the grading will be linear. Let x be the exam score, and let y be the corresponding grade. Find a formula of the form* $y = mx + b$ *which applies to scores x between 40 and 90.*

Exercise 1.2.10 *Find the distance between the pairs of points:*

(a) $(-1,1)$ *and* $(1,1)$.

(b) $(5,3)$ *and* $(-7,-2)$.

(c) $(1,1)$ *and the origin.*

Exercise 1.2.11 *Find the midpoint of the line segment joining the point* $(20,-10)$ *to the origin.*

Exercise 1.2.12 *Find the equation of the circle of radius 3 centered at:*

(a) $(0,0)$

(b) $(5,6)$

(c) $(-5,-6)$

(d) $(0,3)$

(e) $(0,-3)$

(f) $(3,0)$

Exercise 1.2.13 *For each pair of points* $A(x_1,y_1)$ *and* $B(x_2,y_2)$ *find an equation of the circle with center at A that goes through B.*

(a) $A(2,0)$, $B(4,3)$

(b) $A(-2,3)$, $B(4,3)$

Exercise 1.2.14 *Determine the type of conic and sketch it.*

(a) $x^2 + y^2 + 10y = 0$

(b) $9x^2 - 90x + y^2 + 81 = 0$

(c) $6x + y^2 - 8y = 0$

Exercise 1.2.15 *Find the standard equation of the circle passing through* $(-2,1)$ *and tangent to the line* $3x - 2y = 6$ *at the point* $(4,3)$*. Sketch. (Hint: The line through the center of the circle and the point of tangency is perpendicular to the tangent line.)*

1.3 Trigonometry

In this section we review the definitions of trigonometric functions.

1.3.1. Angles and Sectors of Circles

Mathematicians tend to deal mostly with **radians** and we will see later that some formulas are more elegant when using radians (rather than degrees). The relationship between degrees and radians is:

$$\pi \text{ rad} = 180^\circ.$$

Using this formula, some common angles can be derived:

Degrees	0°	30°	45°	60°	90°	120°	135°	150°	180°	270°	360°
Radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π

Example 1.32: Degrees to Radians

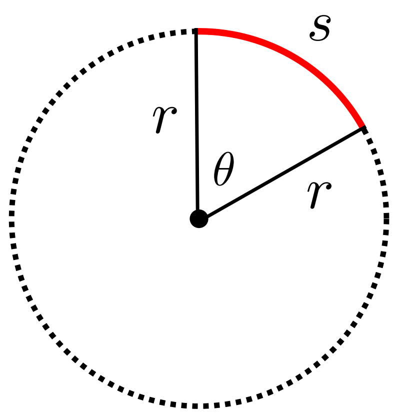
To convert 45° *to radians, multiply by* $\frac{\pi}{180^\circ}$ *to get* $\frac{\pi}{4}$.

Example 1.33: Radians to Degrees

To convert $\frac{5\pi}{6}$ *radians to degrees, multiply by* $\frac{180^\circ}{\pi}$ *to get* 150°.

From now on, unless otherwise indicated, we will *always* use radian measure.

In the diagram below is a sector of a circle with **central angle** θ and radius r **subtending** an arc with length s.



When θ is measure in radians, we have the following formula relating θ, s and r:

$$\theta = \frac{s}{r} \quad \text{or} \quad s = r\theta.$$

Sector Area

The area of the sector is equal to:

$$\text{Sector Area} = \frac{1}{2}r^2\theta.$$

Example 1.34: Angle Subtended by Arc

If a circle has radius 3 cm, then an angle of 2 rad is subtended by an arc of 6 cm ($s = r\theta = 3 \cdot 2 = 6$).

Example 1.35: Area of Circle

If we substitute $\theta = 2\pi$ (a complete revolution) into the sector area formula we get the area of a circle:

$$A = \frac{1}{2}r^2(2\pi) = \pi r^2$$

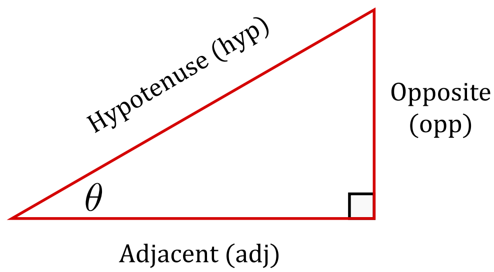
1.3.2. Trigonometric Functions

There are six basic trigonometric functions:

- Sine (abbreviated by sin)
- Cosine (abbreviated by cos)
- Tangent (abbreviated by tan)
- Cosecant (abbreviated by csc)

- Secant (abbreviated by sec)
- Cotangent (abbreviated by cot)

We first describe trigonometric functions in terms of ratios of two sides of a *right angle triangle* containing the angle θ.



With reference to the above triangle, for an acute angle θ (that is, $0 \le \theta < \pi/2$), the six trigonometric functions can be described as follows:

$$\sin\theta = \frac{\text{opp}}{\text{hyp}} \qquad \csc\theta = \frac{\text{hyp}}{\text{opp}}$$

$$\cos\theta = \frac{\text{adj}}{\text{hyp}} \qquad \sec\theta = \frac{\text{hyp}}{\text{adj}}$$

$$\tan\theta = \frac{\text{opp}}{\text{adj}} \qquad \cot\theta = \frac{\text{adj}}{\text{opp}}$$

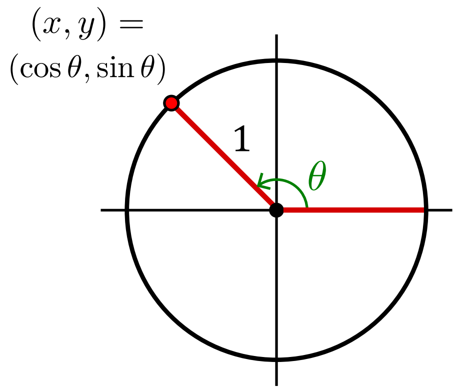
Notice that sin is the ratio of the **o**pposite and **h**ypotenuse. We use the mnemonic SOH to remember this ratio. Similarly, CAH and TOA remind us of the cos and tan ratios.

Mnemonic

The mnemonic *SOH CAH TOA* is useful in remembering how trigonometric functions of acute angles relate to the sides of a right triangle.

This description does not apply to *obtuse* or *negative angles*. To define the six basic trigonometric functions we first define sine and cosine as the lengths of various line segments from a unit circle, and then we define the remaining four basic trigonometric functions in terms of sine and cosine.

Take a line originating at the origin (making an angle of θ with the positive half of the x-axis) and suppose this line intersects the unit circle at the point (x,y). The x- and y-coordinates of this point of intersection are equal to $\cos\theta$ and $\sin\theta$, respectively.



For angles greater than 2π or less than -2π, simply continue to rotate around the circle. In this way, sine and cosine become periodic functions with period 2π:

$$\sin\theta = \sin(\theta + 2\pi k) \qquad\qquad \cos\theta = \cos(\theta + 2\pi k)$$

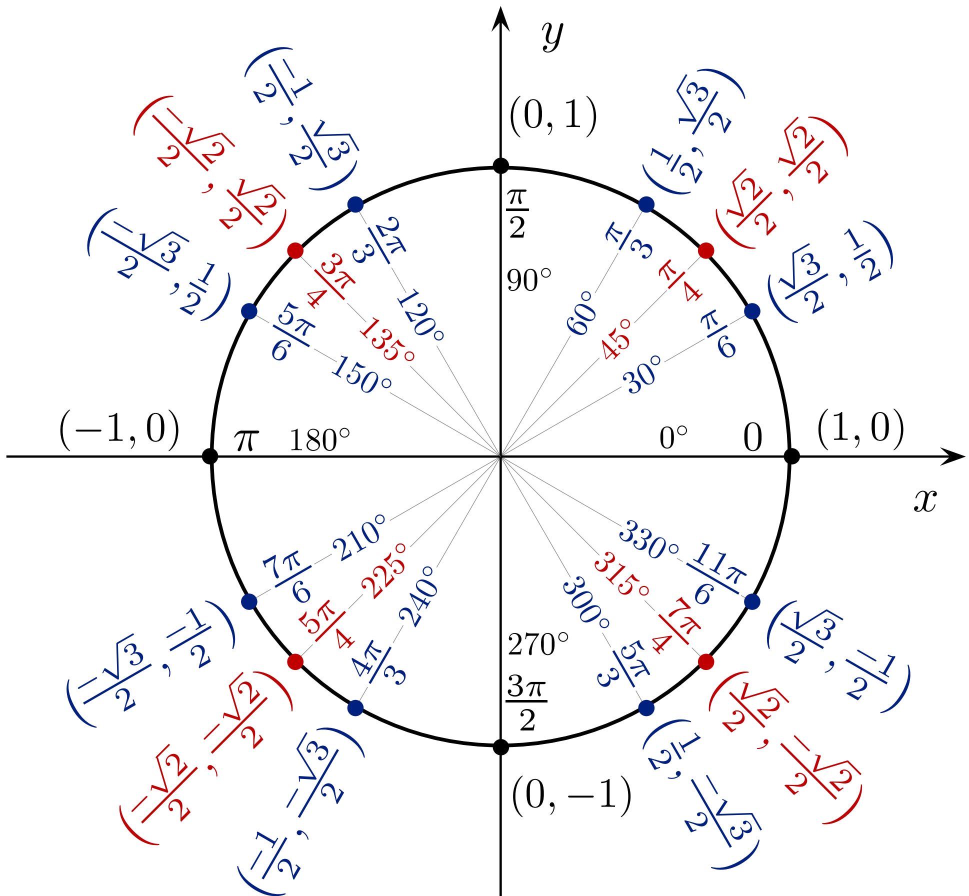
for any angle θ and any integer k.

Above, only sine and cosine were defined directly by the circle. We now define the remaining four basic trigonometric functions in terms of the functions $\sin\theta$ and $\cos\theta$:

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

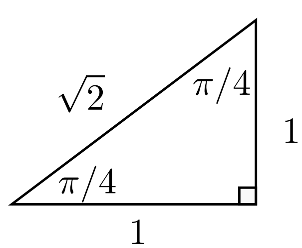
1.3.3. Computing Exact Trigonometric Ratios

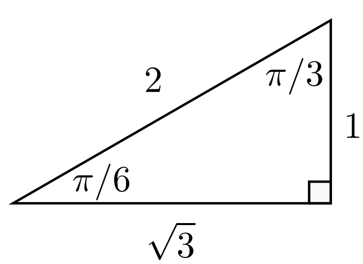
The **unit circle** is often used to determine the *exact* value of a particular trigonometric function.



Reading from the unit circle one can see that $\cos 5\pi/6 = -\sqrt{3}/2$ and $\sin 5\pi/6 = 1/2$ (remember the that the x-coordinate is $\cos\theta$ and the y-coordinate is $\sin\theta$). However, we don't always have access to the unit circle. In this case, we can compute the exact trigonometric ratios for $\theta = 5\pi/6$ by using **special triangles** and the **CAST rule** described below.

The first special triangle has angles of $45^\circ, 45^\circ, 90^\circ$ (i.e., $\pi/4, \pi/4, \pi/2$) with side lengths $1, 1, \sqrt{2}$, while the second special triangle has angles of $30^\circ, 60^\circ, 90^\circ$ (i.e., $\pi/6, \pi/3, \pi/2$) with side lengths $1, 2, \sqrt{3}$. They are classically referred to as the 1-1-$\sqrt{2}$ triangle, and the 1-2-$\sqrt{3}$ triangle, respectively, shown below.





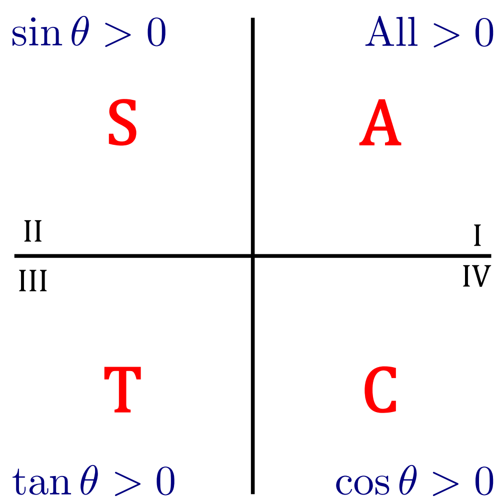
Mnemonic

The first triangle should be easy to remember. To remember the second triangle, place the largest number (2) across from the largest angle ($90^\circ = \pi/2$). Place the smallest number (1) across from the smallest angle ($30^\circ = \pi/6$). Place the middle number ($\sqrt{3} \approx 1.73$) across from the middle angle ($60^\circ = \pi/3$). Double check using the Pythagorean Theorem that the sides satisfy $a^2 + b^2 = c^2$.

The special triangles allow us to compute the exact value (excluding the sign) of trigonometric ratios, but to determine the sign, we can use the *CAST rule*.

The CAST Rule

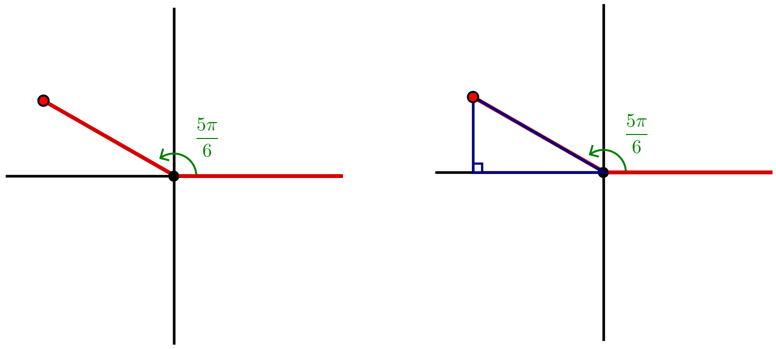
The CAST rule says that in quadrant I all three of $\sin\theta$, $\cos\theta$, $\tan\theta$ are positive. In quadrant II, only $\sin\theta$ is positive, while $\cos\theta$, $\tan\theta$ are negative. In quadrant III, only $\tan\theta$ is positive, while $\sin\theta$, $\cos\theta$ are negative. In quadrant IV, only $\cos\theta$ is positive, while $\sin\theta$, $\tan\theta$ are negative. To remember this, simply label the quadrants by the letters C-A-S-T starting in the bottom right and labelling counter-clockwise.



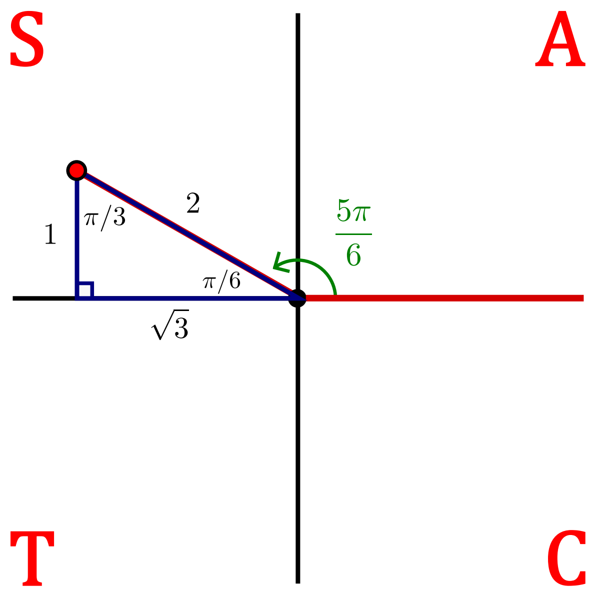
Example 1.36: Determining Trigonometric Ratios Without Unit Circle

Determine $\sin 5\pi/6$, $\cos 5\pi/6$, $\tan 5\pi/6$, $\sec 5\pi/6$, $\csc 5\pi/6$ *and* $\cot 5\pi/6$ *exactly by using the special triangles and CAST rule.*

Solution. We start by drawing the xy-plane and indicating our angle of $5\pi/6$ in standard position (positive angles rotate *counterclockwise* while negative angles rotate *clockwise*). Next, we drop a perpendicular to the x-axis (never drop it to the y-axis!).



Notice that we can now figure out the angles in the triangle. Since $180^\circ = \pi$, we have an interior angle of $\pi - 5\pi/6 = \pi/6$ inside the triangle. As the *angles of a triangle add up to* $180^\circ = \pi$, the other angle must be $\pi/3$. This gives one of our special triangles. We label it accordingly and add the CAST rule to our diagram.



From the above figure we see that $5\pi/6$ lies in quadrant II where $\sin\theta$ is positive and $\cos\theta$ and $\tan\theta$ are negative. This gives us the *sign* of $\sin\theta$, $\cos\theta$ and $\tan\theta$. To determine the *value* we use the special triangle and SOH CAH TOA.

Using $\sin\theta = opp/hyp$ we find a value of $1/2$. Since $\sin\theta$ is positive in quadrant II, we have

$$\sin\frac{5\pi}{6} = +\frac{1}{2}.$$

Using $\cos\theta = adj/hyp$ we find a value of $\sqrt{3}/2$. But $\cos\theta$ is negative in quadrant II, therefore,

$$\cos\frac{5\pi}{6} = -\frac{\sqrt{3}}{2}.$$

Using $\tan\theta = opp/adj$ we find a value of $1/\sqrt{3}$. But $\tan\theta$ is negative in quadrant II, therefore,

$$\tan\frac{5\pi}{6} = -\frac{1}{\sqrt{3}}.$$

To determine $\sec\theta$, $\csc\theta$ and $\cot\theta$ we use the definitions:

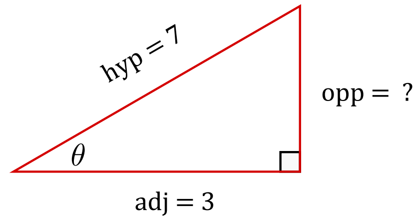
$$\csc\frac{5\pi}{6} = \frac{1}{\sin\frac{5\pi}{6}} = +2, \qquad \sec\frac{5\pi}{6} = \frac{1}{\cos\frac{5\pi}{6}} = -\frac{2}{\sqrt{3}}, \qquad \cot\frac{5\pi}{6} = \frac{1}{\tan\frac{5\pi}{6}} = -\sqrt{3}.$$

♣

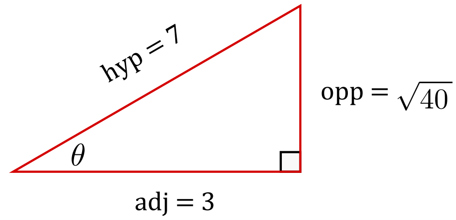
Example 1.37: CAST Rule

If $\cos\theta = 3/7$ and $3\pi/2 < \theta < 2\pi$, then find $\cot\theta$.

Solution. We first draw a right angle triangle. Since $\cos\theta = adj/hyp = 3/7$, we let the adjacent side have length 3 and the hypotenuse have length 7.



Using the Pythagorean Theorem, we have $3^2 + (\text{opp})^2 = 7^2$. Thus, the opposite side has length $\sqrt{40}$.



To find $\cot\theta$ we use the definition:

$$\cot\theta = \frac{1}{\tan\theta}.$$

Since we are given $3\pi/2 < \theta < 2\pi$, we are in the fourth quadrant. By the CAST rule, $\tan\theta$ is negative in this quadrant. As $\tan\theta = opp/adj$, it has a value of $\sqrt{40}/3$, but by the CAST rule it is negative, that is,

$$\tan\theta = -\frac{\sqrt{40}}{3}.$$

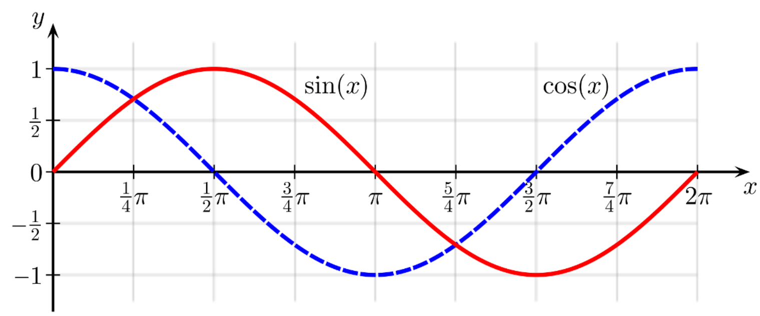
Therefore,

$$\cot\theta = -\frac{3}{\sqrt{40}}.$$

♣

1.3.4. Graphs of Trigonometric Functions

The graph of the functions $\sin x$ and $\cos x$ can be visually represented as:

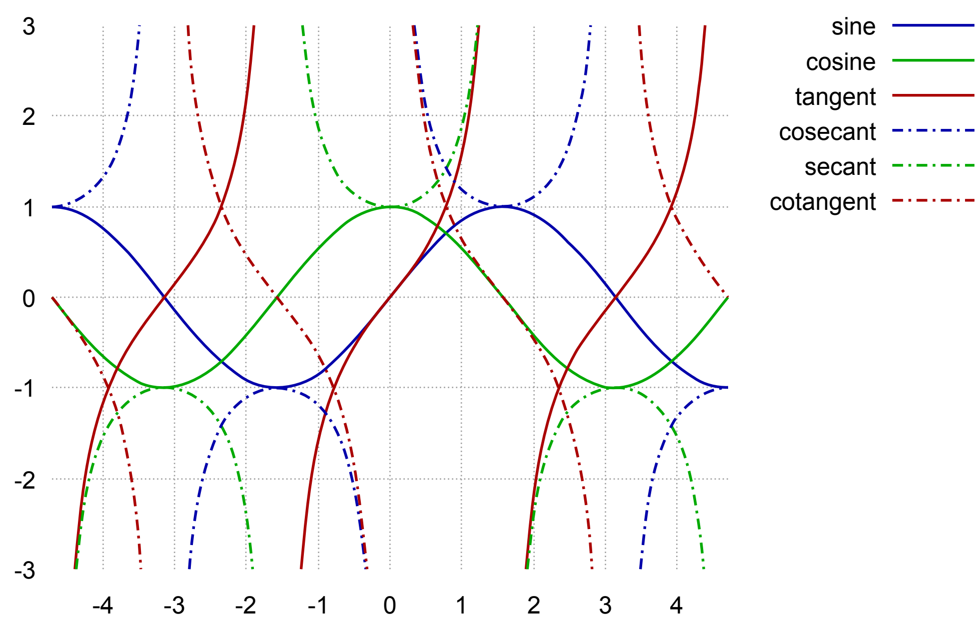


Both $\sin x$ and $\cos x$ have domain $(-\infty, \infty)$ and range $[-1, 1]$. That is,

$$-1 \leq \sin x \leq 1 \qquad -1 \leq \cos x \leq 1.$$

The zeros of $\sin x$ occur at the integer multiples of π, that is, $\sin x = 0$ whenever $x = n\pi$, where n is an integer. Similarly, $\cos x = 0$ whenever $x = \pi/2 + n\pi$, where n is an integer.

The six basic trigonometric functions can be visually represented as:



Both tangent and cotangent have range $(-\infty, \infty)$, whereas cosecant and secant have range $(-\infty, -1] \cup [1, \infty)$. Each of these functions is periodic. Tangent and cotangent have period π, whereas sine, cosine, cosecant and secant have period 2π.

1.3.5. Trigonometric Identities

There are numerous trigonometric identities, including those relating to shift/periodicity, Pythagoras type identities, double-angle formulas, half-angle formulas and addition formulas. We list these below.

1. **Shifts and periodicity**

$\sin(\theta+2\pi) = \sin\theta$	$\cos(\theta+2\pi) = \cos\theta$	$\tan(\theta+2\pi) = \tan\theta$
$\sin(\theta+\pi) = -\sin\theta$	$\cos(\theta+\pi) = -\cos\theta$	$\tan(\theta+\pi) = \tan\theta$
$\sin(-\theta) = -\sin\theta$	$\cos(-\theta) = \cos\theta$	$\tan(-\theta) = -\tan\theta$
$\sin\left(\frac{\pi}{2}-\theta\right) = \cos\theta$	$\cos\left(\frac{\pi}{2}-\theta\right) = \sin\theta$	$\tan\left(\frac{\pi}{2}-\theta\right) = \cot\theta$

2. **Pythagoras type formulas**
   - $\sin^2\theta+\cos^2\theta = 1$
   - $\tan^2\theta+1 = \sec^2\theta$
   - $1+\cot^2\theta = \csc^2\theta$

3. **Double-angle formulas**
   - $\sin(2\theta) = 2\sin\theta\cos\theta$
   - $$\begin{aligned}\cos(2\theta) &= \cos^2\theta-\sin^2\theta\\ &= 2\cos^2\theta-1\\ &= 1-2\sin^2\theta.\end{aligned}$$

4. **Half-angle formulas**
   - $\cos^2\theta = \dfrac{1+\cos(2\theta)}{2}$
   - $\sin^2\theta = \dfrac{1-\cos(2\theta)}{2}$

5. **Addition formulas**
   - $\sin(\theta+\phi) = \sin\theta\cos\phi+\cos\theta\sin\phi$
   - $\cos(\theta+\phi) = \cos\theta\cos\phi-\sin\theta\sin\phi$
   - $\tan(\theta+\phi) = \dfrac{\tan\theta+\tan\phi}{1-\tan\theta\tan\phi}$
   - $\sin(\theta-\phi) = \sin\theta\cos\phi-\cos\theta\sin\phi$
   - $\cos(\theta-\phi) = \cos\theta\cos\phi+\sin\theta\sin\phi$

Example 1.38: Double Angle

Find all values of x with $0 \le x \le \pi$ such that $\sin 2x = \sin x$.

Solution. Using the double-angle formula $\sin 2x = 2\sin x\cos x$ we have:

$$2\sin x\cos x = \sin x$$

$$2\sin x\cos x - \sin x = 0$$

$$\sin x(2\cos x-1) = 0$$

Thus, either $\sin x = 0$ or $\cos x = 1/2$. For the first case when $\sin x = 0$, we get $x = 0$ or $x = \pi$. For the second case when $\cos x = 1/2$, we get $x = \pi/3$ (use the special triangles and CAST rule to get this). Thus, we have three solutions: $x = 0$, $x = \pi/3$, $x = \pi$. ♣

Exercises for 1.3

Exercise 1.3.1 *Find all values of* θ *such that* $\sin(\theta) = -1$*; give your answer in radians.*

Exercise 1.3.2 *Find all values of* θ *such that* $\cos(2\theta) = 1/2$*; give your answer in radians.*

Exercise 1.3.3 *Compute the following:*

(a) $\sin(3\pi)$

(b) $\sec(5\pi/6)$

(c) $\cos(-\pi/3)$

(d) $\csc(4\pi/3)$

(e) $\tan(7\pi/4)$

(f) $\cot(13\pi/4)$

Exercise 1.3.4 *If* $\sin\theta = \frac{3}{5}$ *and* $\frac{\pi}{2} < \theta < \pi$*, then find* $\sec\theta$.

Exercise 1.3.5 *Suppose that* $\tan\theta = x$ *and* $\pi < \theta < \frac{3\pi}{2}$*, find* $\sin\theta$ *and* $\cos\theta$ *in terms of* x.

Exercise 1.3.6 *Find an angle* θ *such that* $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ *and* $\sin\theta = \sin\frac{23\pi}{7}$.

Exercise 1.3.7 *Use an angle sum identity to compute* $\cos(\pi/12)$.

Exercise 1.3.8 *Use an angle sum identity to compute* $\tan(5\pi/12)$.

Exercise 1.3.9 *Verify the following identities*

(a) $\cos^2(t)/(1-\sin(t)) = 1+\sin(t)$

(b) $2\csc(2\theta) = \sec(\theta)\csc(\theta)$

(c) $\sin(3\theta) - \sin(\theta) = 2\cos(2\theta)\sin(\theta)$

Exercise 1.3.10 *Sketch the following functions:*

(a) $y = 2\sin(x)$

(b) $y = \sin(3x)$

(c) $y = \sin(-x)$

Exercise 1.3.11 *Find all of the solutions of* $2\sin(t) - 1 - \sin^2(t) = 0$ *in the interval* $[0, 2\pi]$.

1.4 Additional Exercises

These problems require a comprehensive knowledge of the skills reviewed in this chapter. They are not in any particular order. A proficiency in these skills will help you a long way as your learn the calculus material in the following chapters.

Exercise 1.4.1 *Rationalize the denominator for each of the following expressions. That is, re-write the expression in such a way that no square roots appear in the denominator. Also, simplify your answers if possible.*

(a) $\dfrac{1}{\sqrt{2}}$

(b) $\dfrac{3h}{\sqrt{x+h+1}-\sqrt{x+1}}$

Exercise 1.4.2 *Solve the following equations.*

(a) $2-5(x-3)=4-10x$

(b) $2x^2-5x=3$

(c) $x^2-x-3=0$

(d) $x^2+x+3=0$

(e) $\sqrt{x^2+9}=2x$

Exercise 1.4.3 *By means of counter-examples, show why it is wrong to say that the following equations hold for all real numbers for which the expressions are defined.*

(a) $(x-2)^2=x^2-2^2$

(b) $\dfrac{1}{x+h}=\dfrac{1}{x}+\dfrac{1}{h}$

(c) $\sqrt{x^2+y^2}=x+y$

Exercise 1.4.4 *Find an equation of the line passing through the point* $(-2,5)$ *and parallel to the line* $x+3y-2=0$.

Exercise 1.4.5 *Solve* $\dfrac{x^2-1}{3x-1}\leq 1$.

Exercise 1.4.6 *Explain why the following expression never represents a real number (for any real number* x*):* $\sqrt{x-2}+\sqrt{1-x}$.

Exercise 1.4.7 *Simplify the expression* $\dfrac{\left[3(x+h)^2+4\right]-\left[3x^2+4\right]}{h}$ *as much as possible.*

Exercise 1.4.8 *Simplify the expression* $\dfrac{\frac{x+h}{2(x+h)-1}-\frac{x}{2x-1}}{h}$ *as much as possible.*

Exercise 1.4.9 *Simplify the expression* $-\sin x(\cos x+3\sin x)-\cos x(-\sin x+3\cos x)$.

Exercise 1.4.10 *Solve the equation* $\cos x=\frac{\sqrt{3}}{2}$ *on the interval* $0\leq x\leq 2\pi$.

Exercise 1.4.11 *Find an angle* θ *such that* $0\leq\theta\leq\pi$ *and* $\cos\theta=\cos\frac{38\pi}{5}$.

Exercise 1.4.12 *What can you say about* $\dfrac{|x|+|4-x|}{x-2}$ *when* x *is a large (positive) number?*

Exercise 1.4.13 *Find an equation of the circle with centre in* $(-2,3)$ *and passing through the point* $(1,-1)$.

Exercise 1.4.14 *Find the centre and radius of the circle described by* $x^2+y^2+6x-4y+12=3$.

Exercise 1.4.15 *If* $y=9x^2+6x+7$, *find all possible values of* y.

Exercise 1.4.16 *Simplify* $\left(\dfrac{3x^2y^3z^{-1}}{18x^{-1}yz^3}\right)^2$.

Exercise 1.4.17 *If* $y=\dfrac{3x+2}{1-4x}$, *then what is* x *in terms of* y*?*

Exercise 1.4.18 *Divide* x^2+3x-5 *by* $x+2$ *to obtain the quotient and the remainder. Equivalently, find polynomial* $Q(x)$ *and constant* R *such that*

$$\frac{x^2+3x-5}{x+2}=Q(x)+\frac{R}{x+2}.$$

2. Functions

2.1 What is a Function?

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

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

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

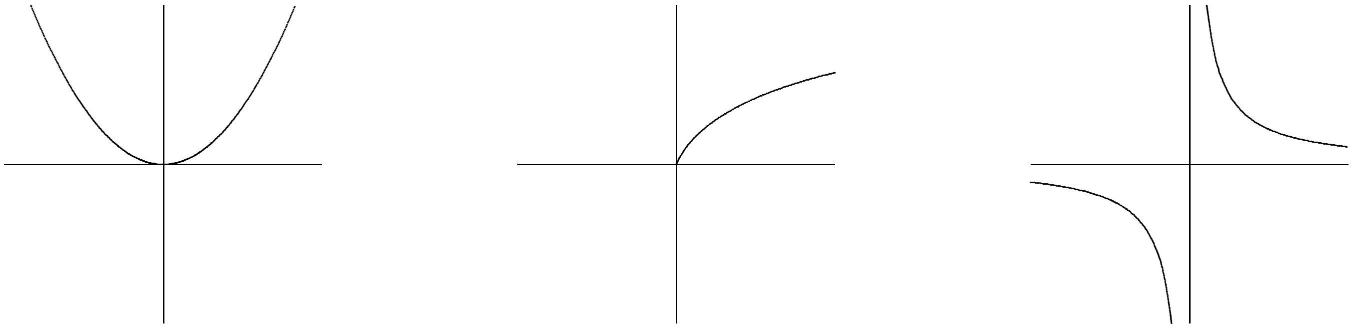
One test to identify whether or not a curve in the (x, y) coordinate system is a function is the following.

Theorem 2.1: The Vertical Line Test

A curve in the (x, y) coordinate system represents a function if and only if no vertical line intersects the curve more than once.

In addition to lines, another familiar example of a function is the parabola $y = f(x) = x^2$. We can draw the graph of this function by taking various values of x (say, at regular intervals) and plotting the points $(x, f(x)) = (x, x^2)$. Then connect the points with a smooth curve. (See figure 2.1.)

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



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

Figure 2.1: Graphs of Functions

Example 2.2: Domain of the Square-Root Function

The square-root function $y = f(x) = \sqrt{x}$ is the rule which says, given an x-value, take the nonnegative number whose square is x. This rule only makes sense if $x \geq 0$. We say that the domain of this function is $x \geq 0$, or more formally $\{x \in \mathbb{R} : x \geq 0\}$. Alternately, we can use interval notation, and write that the domain is $[0, \infty)$. The fact that the domain of $y = \sqrt{x}$ is $[0, \infty)$ means that in the graph of this function (see figure 2.1) we have points (x, y) only above x-values on the right side of the x-axis.

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

To summarize, two reasons why certain x-values are excluded from the domain of a function are the following.

Restrictions for the Domain of a Function

1. We cannot divide by zero, and

2. We cannot take the square root of a negative number.

When the domain of a function is restricted, we say that the function is undefined at that point. We will encounter some other ways in which functions might be undefined later.

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

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

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

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

Example 2.3: Open Box

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

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

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

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

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

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

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

Example 2.4: Circle of Radius r Centered at the Origin

Is the circle of radius r centered at the origin the graph of a function?

Solution. The equation for this circle is usually given in the form $x^2 + y^2 = r^2$. To write the equation in the form $y = f(x)$ we solve for y, obtaining $y = \pm\sqrt{r^2 - x^2}$. But *this is not a function*, because when we substitute a value in the interval $(-r, r)$ for x there are two corresponding values of y. To get a function, we must choose one of the two signs in front of the square root. If we choose the positive sign, for example, we get the upper semicircle $y = f(x) = \sqrt{r^2 - x^2}$ (see figure 2.2). The domain of this function is the interval $[-r, r]$, i.e., x must be between $-r$ and r (including the endpoints). If x is outside of that interval, then $r^2 - x^2$ is negative, and we cannot take the square root. In terms of the graph, this just means that there are no points on the curve whose x-coordinate is greater than r or less than $-r$. ♣

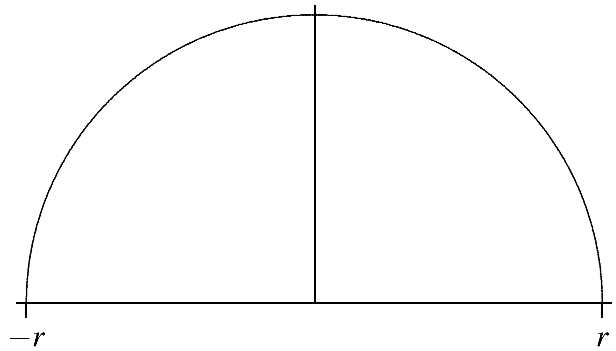


Figure 2.2: Upper semicircle $y = \sqrt{r^2 - x^2}$.

Example 2.5: Domain

Find the domain of

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

Solution. To answer this question, we must rule out the x-values that make $4x - x^2$ negative (because we cannot take the square root of a negative number) and also the x-values that make $4x - x^2$ zero (because if $4x - x^2 = 0$, then when we take the square root we get 0, and we cannot divide by 0). In other words, the domain consists of all x for which $4x - x^2$ is strictly positive. The inequality $4x - x^2 > 0$ was solved in Example 1.12. In interval notation, the domain is the interval $(0, 4)$. ♣

A function does not always have to be given by a single formula as the next example demonstrates.

Example 2.6: Piecewise Velocity

Suppose that $y = v(t)$ is the velocity function for a car which starts out from rest (zero velocity) at time $t = 0$; then increases its speed steadily to 20 m/sec, taking 10 seconds to do this; then travels at constant speed 20 m/sec for 15 seconds; and finally applies the brakes to decrease speed steadily to 0, taking 5 seconds to do this. The formula for $y = v(t)$ is different in each of the three time intervals: first $y = 2x$, then $y = 20$, then $y = -4x + 120$. The graph of this function is shown in figure 2.3.

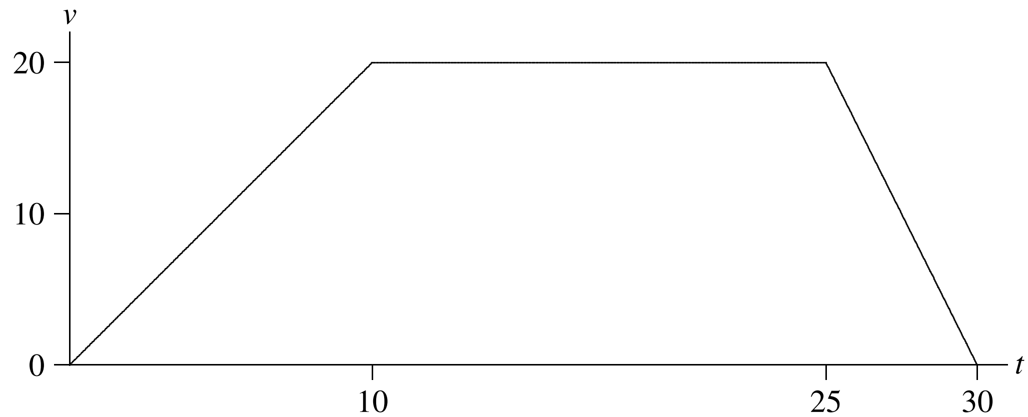


Figure 2.3: A velocity function.

Our focus in this textbook will be on **elementary functions**. Examples of elementary functions include both *algebraic* functions, such as polynomials, rational functions, and powers, and *transcendental* functions, such as exponential, logarithmic, and trigonometric functions.

Exercises for 2.1

Exercise 2.1.1 *Find the domain of each of the following functions:*

(a) $y = x^2 + 1$

(b) $y = f(x) = \sqrt{2x-3}$

(c) $y = f(x) = 1/(x+1)$

(d) $y = f(x) = 1/(x^2-1)$

(e) $y = f(x) = \sqrt{-1/x}$

(f) $y = f(x) = \sqrt[3]{x}$

(g) $y = f(x) = \sqrt{r^2-(x-h)^2}$, *where* r *and* h *are positive constants.*

(h) $y = f(x) = \sqrt[4]{x}$

(i) $y = \sqrt{1-x^2}$

(j) $y = f(x) = \sqrt{1-(1/x)}$

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

(l) $y = f(x) = \sqrt{x} + 1/(x-1)$

(m) $y = f(x) = 1/(\sqrt{x}-1)$

Exercise 2.1.2 *A farmer wants to build a fence along a river. He has 500 feet of fencing and wants to enclose a rectangular pen on three sides (with the river providing the fourth side). If* x *is the length of the side perpendicular to the river, determine the area of the pen as a function of* x*. What is the domain of this function?*

Exercise 2.1.3 *A can in the shape of a cylinder is to be made with a total of 100 square centimeters of material in the side, top, and bottom; the manufacturer wants the can to hold the maximum possible volume. Write the volume as a function of the radius* r *of the can; find the domain of the function.*

Exercise 2.1.4 *A can in the shape of a cylinder is to be made to hold a volume of one liter (1000 cubic centimeters). The manufacturer wants to use the least possible material for the can. Write the surface area of the can (total of the top, bottom, and side) as a function of the radius r of the can; find the domain of the function.*

2.2 Transformations and Compositions

2.2.1. Transformations

Transformations are operations we can apply to a function in order to obtain a *new* function. The most common transformations include translations (shifts), stretches and reflections. We summarize these below.

Function	Conditions	How to graph $F(x)$ given the graph of $f(x)$
$F(x) = f(x) + c$	$c > 0$	Shift $f(x)$ upwards by c units
$F(x) = f(x) - c$	$c > 0$	Shift $f(x)$ downwards by c units
$F(x) = f(x + c)$	$c > 0$	Shift $f(x)$ to the left by c units
$F(x) = f(x - c)$	$c > 0$	Shift $f(x)$ to the right by c units
$F(x) = -f(x)$		Reflect $f(x)$ about the x-axis
$F(x) = f(-x)$		Reflect $f(x)$ about the y-axis
$F(x) = \|f(x)\|$		Take the part of the graph of $f(x)$ that lies below the x-axis and reflect it about the x-axis

For horizontal and vertical stretches, different resources use different terminology and notation. Use the one you are most comfortable with! Below, both a, b are positive numbers. Note that we only use the term *stretch* in this case:

Function	Conditions	How to graph $F(x)$ given the graph of $f(x)$
$F(x) = af(x)$	$a > 0$	Stretch $f(x)$ vertically by a factor of a
$F(x) = f(bx)$	$b > 0$	Stretch $f(x)$ horizontally by a factor of $1/b$

In the next case, we use both the terms *stretch* and *shrink*. We also split up vertical stretches into two cases ($0 < a < 1$ and $a > 1$), and split up horizontal stretches into two cases ($0 < b < 1$ and $b > 1$). Note that having $0 < a < 1$ is the same as having $1/c$ with $c > 1$. Also note that *stretching by a factor of* $1/c$ is the same as *shrinking by a factor* c.

Function	Conditions	How to graph $F(x)$ given the graph of $f(x)$
$F(x) = cf(x)$	$c > 1$	Stretch $f(x)$ vertically by a factor of c
$F(x) = (1/c)f(x)$	$c > 1$	Shrink $f(x)$ vertically by a factor of c
$F(x) = f(cx)$	$c > 1$	Shrink $f(x)$ horizontally by a factor of c
$F(x) = f(x/c)$	$c > 1$	Stretch $f(x)$ horizontally by a factor of c

Some resources keep the condition $0 < c < 1$ rather than using $1/c$. This is illustrated in the next table.

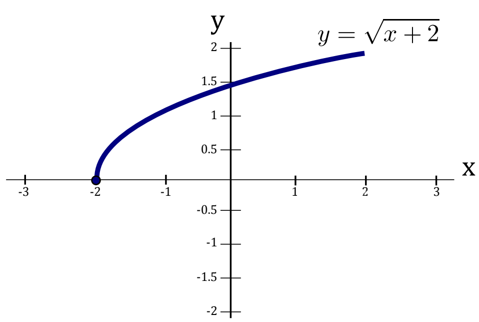
Function	Conditions	How to graph $F(x)$ given the graph of $f(x)$
$F(x) = df(x)$	$d > 1$	Stretch $f(x)$ vertically by a factor of d
$F(x) = df(x)$	$0 < d < 1$	Shrink $f(x)$ vertically by a factor of $1/d$
$F(x) = f(dx)$	$d > 1$	Shrink $f(x)$ horizontally by a factor of d
$F(x) = f(dx)$	$0 < d < 1$	Stretch $f(x)$ horizontally by a factor of $1/d$

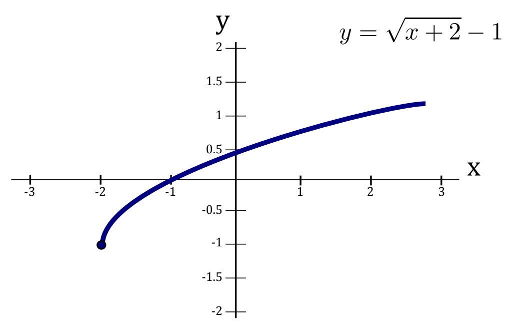
Example 2.7: Transformations and Graph Sketching

In this example we will use appropriate transformations to sketch the graph of the function

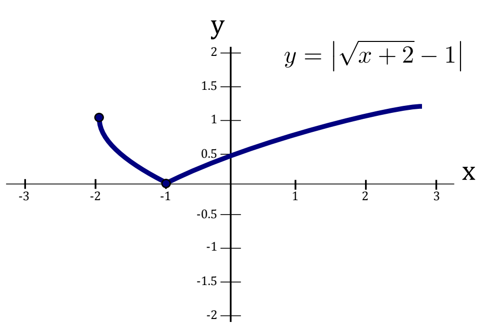
$$y = |\sqrt{x+2} - 1| - 1$$

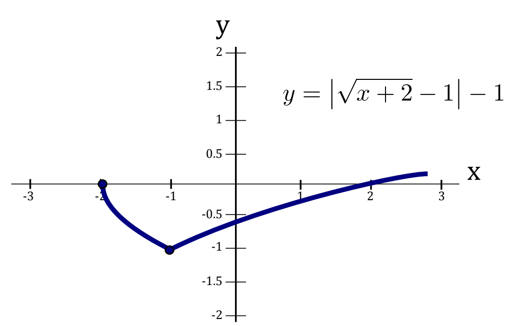
Solution. We start with the graph of a function we know how to sketch, in particular, $y = \sqrt{x}$: To obtain the graph of the function $y = \sqrt{x+2}$ from the graph $y = \sqrt{x}$, we must shift $y = \sqrt{x}$ to the left by 2 units. To obtain the graph of the function $y = \sqrt{x+2} - 1$ from the graph $y = \sqrt{x+2}$, we must shift $y = \sqrt{x+2}$ downwards by 1 unit.





To obtain the graph of the function $y = |\sqrt{x+2} - 1|$ from the graph $y = \sqrt{x+2} - 1$, we must take the part of the graph of $y = \sqrt{x+2} - 1$ that lies below the x-axis and reflect it (upwards) about the x-axis. Finally, to obtain the graph of the function $y = |\sqrt{x+2} - 1| - 1$ from the graph $y = |\sqrt{x+2} - 1|$, we must shift $y = |\sqrt{x+2} - 1|$ downwards by 1 unit:





♣

2.2.2. Combining Two Functions

Let f and g be two functions. Then we can form new functions by adding, subtracting, multiplying, or dividing. These new functions, $f+g$, $f-g$, fg and f/g, are defined in the usual way.

Operations on Functions

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

$$(fg)(x) = f(x)g(x) \qquad \left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$$

Suppose D_f is the domain of f and D_g is the domain of g. Then the domains of $f+g$, $f-g$ and fg are the same and are equal to the intersection $D_f \cap D_g$ (that is, everything that is in *common* to both the domain of f and the domain of g). Since division by zero is *not allowed*, the domain of f/g is $\{x \in D_f \cap D_g : g(x) \neq 0\}$.

Another way to combine two functions f and g together is a procedure called composition.

Function Composition

Given two functions f and g, the **composition** of f and g, denoted by $f \circ g$, is defined as:

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

The domain of $f \circ g$ is $\{x \in D_g : g(x) \in D_f\}$, that is, it contains all values x in the domain of g such that $g(x)$ is in the domain of f.

Example 2.8: Domain of a Composition

Let $f(x) = x^2$ and $g(x) = \sqrt{x}$. Find the domain of $f \circ g$.

Solution. The domain of f is $D_f = \{x \in \mathbb{R}\}$. The domain of g is $D_g = \{x \in \mathbb{R} : x \geq 0\}$. The function $(f \circ g)(x) = f(g(x))$ is:

$$f(g(x)) = \left(\sqrt{x}\right)^2 = x.$$

Typically, $h(x) = x$ would have a domain of $\{x \in \mathbb{R}\}$, but since it came from a **composed function**, we must consider $g(x)$ when looking at the domain of $f(g(x))$. Thus, the domain of $f \circ g$ is $\{x \in \mathbb{R} : x \geq 0\}$.

♣

Example 2.9: Combining Two Functions

Let $f(x) = x^2+3$ and $g(x) = x-2$. Find $f+g$, $f-g$, fg, f/g, $f \circ g$ and $g \circ f$. Also, determine the domains of these new functions.

Solution. For $f+g$ we have:

$$(f+g)(x) = f(x)+g(x) = (x^2+3)+(x-2) = x^2+x+1.$$

For $f-g$ we have:

$$(f-g)(x) = f(x)-g(x) = (x^2+3)-(x-2) = x^2+3-x+2 = x^2-x+5.$$

For fg we have:

$$(fg)(x) = f(x)\cdot g(x) = (x^2+3)(x-2) = x^3-2x^2+3x-6.$$

For f/g we have:

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

For $f\circ g$ we have:

$$(f\circ g)(x) = f(g(x)) = f(x-2) = (x-2)^2+3 = x^2-4x+7.$$

For $g\circ f$ we have:

$$(g\circ f)(x) = g(f(x)) = g(x^2+3) = (x^2+3)-2 = x^2+1.$$

The domains of $f+g$, $f-g$, fg, $f\circ g$ and $g\circ f$ is $\{x\in\mathbb{R}\}$, while the domain of f/g is $\{x\in\mathbb{R} : x\neq 2\}$.
♣

As in the above problem, $f\circ g$ and $g\circ f$ are generally different functions.

Exercises for 2.2

Exercise 2.2.1 *Starting with the graph of $y=\sqrt{x}$, the graph of $y=1/x$, and the graph of $y=\sqrt{1-x^2}$ (the upper unit semicircle), sketch the graph of each of the following functions:*

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

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

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

(d) $y=f(x)=x/(1-x)$

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

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

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

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

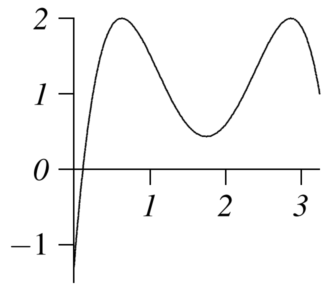
(i) $f(x)=1/(x+1)$

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

(k) $f(x)=1+1/(x-1)$

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

Exercise 2.2.2 *The graph of $f(x)$ is shown below. Sketch the graphs of the following functions.*



(a) $y = f(x-1)$

(b) $y = 1 + f(x+2)$

(c) $y = 1 + 2f(x)$

(d) $y = 2f(3x)$

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

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

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

(h) $y = |f(x) - 2|$

Exercise 2.2.3 *Suppose $f(x) = 3x - 9$ and $g(x) = \sqrt{x}$. What is the domain of the composition $(g \circ f)(x)$?*

2.3 Exponential Functions

An **exponential function** is a function of the form $f(x) = a^x$, where a is a constant. Examples are 2^x, 10^x and $(1/2)^x$. To more formally define the exponential function we look at various kinds of input values.

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

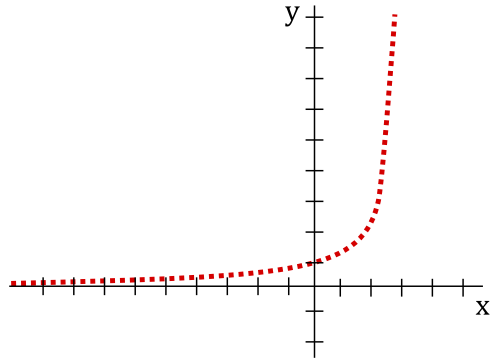
We'll evade the full answer to the hard question, but we have to know something about exponential functions. You need first to understand that since it's not "obvious" what 2^x should mean, we are really free to make it mean whatever we want, so long as we keep the behavior that *is* obvious, namely,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

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

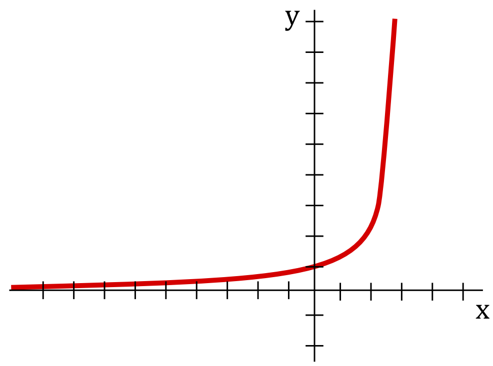
The next easiest set of numbers to understand is the negative integers: for example, $2^{-3} = 1/2^3$. We know that whatever 2^{-3} means it must be that $2^{-3}2^3 = 2^{-3+3} = 2^0 = 1$, which means that 2^{-3} must be $1/2^3$. In fact, by the same argument, once we know what 2^x means for some value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

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

What's left is the hard part: what does 2^x mean when x cannot be written as a fraction, like $x = \sqrt{2}$ or $x = \pi$? What we know so far is how to assign meaning to 2^x whenever $x = p/q$. If we were to graph a^x (for some $a > 1$) at points $x = p/q$ then we'd see something like this:

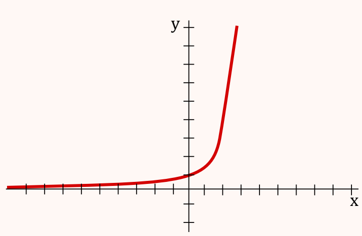


This is a poor picture, but it illustrates a series of individual points above the rational numbers on the x-axis. There are really a lot of "holes" in the curve, above $x = \pi$, for example. But (this is the hard part) it is possible to prove that the holes can be "filled in", and that the resulting function, called a^x, really does have the properties we want, namely that $a^x a^y = a^{x+y}$ and $(a^x)^y = a^{xy}$. Such a graph would then look like this:

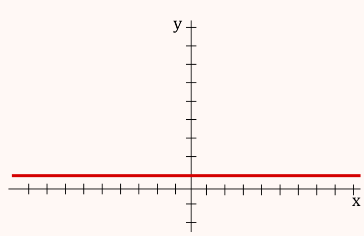


Three Types of Exponential Functions

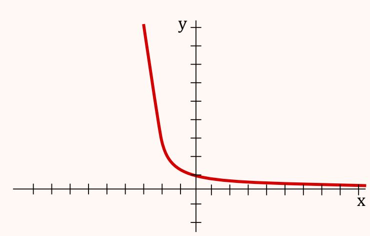
There are *three kinds* of exponential functions $f(x) = a^x$ depending on whether $a > 1$, $a = 1$ or $0 < a < 1$:



$f(x) = a^x$
$a > 1$



$f(x) = 1^x$



$f(x) = a^x$
$0 < a < 1$

Properties of Exponential Functions

The first thing to note is that if $a < 0$ then problems can occur. Observe that if $a = -1$ then $(-1)^x$ is not defined for every x. For example, $x = 1/2$ is a square root and gives $(-1)^{1/2} = \sqrt{-1}$ which is not a real number.

Exponential Function Properties

- *Only defined for positive a:* a^x is only defined for all real x if $a > 0$
- *Always positive:* $a^x > 0$, for all x
- *Exponent rules:*
  1. $a^x a^y = a^{x+y}$
  2. $\dfrac{a^x}{a^y} = a^{x-y}$
  3. $(a^x)^y = a^{xy} = a^{yx} = (a^y)^x$
  4. $a^x b^x = (ab)^x$
- *Long term behaviour:* If $a > 1$, then $a^x \to \infty$ as $x \to \infty$ and $a^x \to 0$ as $x \to -\infty$.

The last property can be observed from the graph. If $a > 1$, then as x gets larger and larger, so does a^x. On the other hand, as x gets large and negative, the function approaches the x-axis, that is, a^x approaches 0.

Example 2.10: Reflection of Exponential

Determine an equation of the function after reflecting $y = 2^x$ about the line $x = -2$.

Solution. First reflect about the y-axis to get $y = 2^{-x}$. Now shift by $2 \times 2 = 4$ units to the *left* to get $y = 2^{-(x+4)}$. Side note: Can you see why this sequence of transformations is the same as reflection in the line $x = -2$? Can you come up with a general rule for these types of reflections? ♣

Example 2.11: Determine the Exponential Function

Determine the exponential function $f(x) = ka^x$ that passes through the points $(1,6)$ and $(2,18)$.

Solution. We substitute our two points into the equation to get:

$$x = 1, y = 6 \to 6 = ka^1$$

$$x = 2, y = 18 \to 18 = ka^2$$

This gives us $6 = ka$ and $18 = ka^2$. The first equation is $k = 6/a$ and subbing this into the second gives: $18 = (6/a)a^2$. Thus, $18 = 6a$ and $a = 3$. Now we can see from $6 = ka$ that $k = 2$. Therefore, the exponential function is

$$f(x) = 2 \cdot 3^x.$$

♣

There is one base that is so important and convenient that we give it a special symbol. This number is denoted by $e = 2.71828\ldots$ (and is an irrational number). Its *importance* stems from the fact that it simplifies many formulas of Calculus and also shows up in other fields of mathematics.

Example 2.12: Domain of Function with Exponential

Find the domain of $f(x) = \dfrac{1}{\sqrt{e^x+1}}$.

Solution. For domain, we cannot divide by zero or take the square root of negative numbers. Note that one of the properties of exponentials is that they are always positive! Thus, $e^x + 1 > 0$ (in fact, as $e^x > 0$ we actually have that $e^x + 1$ is at least one). Therefore, $e^x + 1$ is never zero nor negative, and gives no restrictions on x. Thus, the domain is $\mathbb{R}$. ♣

Exercises for 2.3

Exercise 2.3.1 *Determine an equation of the function* $y = a^x$ *passing through the point* $(3,8)$.

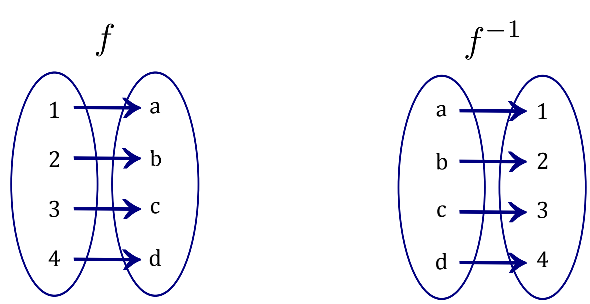
Exercise 2.3.2 *Find the y-intercept of* $f(x) = 4^x + 6$.

Exercise 2.3.3 *Find the y-intercept of* $f(x) = 2\left(\frac{1}{2}\right)^x$.

Exercise 2.3.4 *Find the domain of* $y = e^{-x} + e^{\frac{1}{x}}$.

2.4 Inverse Functions

In mathematics, an *inverse* is a function that serves to "undo" another function. That is, if $f(x)$ produces y, then putting y into the inverse of f produces the output x. A function f that has an inverse is called invertible and the inverse is denoted by f^{-1}. It is best to illustrate inverses using an arrow diagram:



Notice how f maps 1 to a, and f^{-1} undoes this, that is, f^{-1} maps a back to 1. Don't confuse $f^{-1}(x)$ with exponentiation: the inverse f^{-1} is *different* from $\frac{1}{f(x)}$.

Not every function has an inverse. It is easy to see that if a function $f(x)$ is going to have an inverse, then $f(x)$ *never* takes on the same value twice. We give this property a special name.

A function $f(x)$ is called **one-to-one** if every element of the range corresponds to *exactly* one element of the domain. Similar to the Vertical Line Test (VLT) for functions, we have the Horizontal Line Test (HLT) for the one-to-one property.

Theorem 2.13: The Horizontal Line Test

A function is one-to-one if and only if there is no horizontal line that intersects its graph more than once.

Example 2.14: Parabola is Not One-to-one

The parabola $f(x) = x^2$ it not one-to-one because it does not satisfy the horizontal line test. For example, the horizontal line $y = 1$ intersects the parabola at two points, when $x = -1$ and $x = 1$.

We now formally define the inverse of a function.

Definition 2.15: Inverse of a Function

Let $f(x)$ and $g(x)$ be two one-to-one *functions. If $(f \circ g)(x) = x$ and $(g \circ f)(x) = x$ then we say that $f(x)$ and $g(x)$ are* **inverses** *of each other. We denote $g(x)$ (the inverse of $f(x)$) by $g(x) = f^{-1}(x)$.*

Thus, if f maps x to y, then f^{-1} maps y back to x. This gives rise to the *cancellation formulas*:

$$f^{-1}(f(x)) = x, \quad \text{for every } x \text{ in the domain of } f(x),$$

$$f(f^{-1}(x)) = x, \quad \text{for every } x \text{ in the domain of } f^{-1}(x).$$

Example 2.16: Finding the Inverse at Specific Values

If $f(x) = x^9 + 2x^7 + x + 1$, find $f^{-1}(5)$ and $f^{-1}(1)$.

Solution. Rather than trying to compute a formula for f^{-1} and then computing $f^{-1}(5)$, we can simply find a number c such that f evaluated at c gives 5. Note that subbing in some simple values ($x = -3, -2, 1, 0, 1, 2, 3$) and evaluating $f(x)$ we eventually find that $f(1) = 1^9 + 2(1^7) + 1 + 1 = 5$ and $f(0) = 1$. Therefore, $f^{-1}(5) = 1$ and $f^{-1}(1) = 0$. ♣

To compute the equation of the inverse of a function we use the following *guidelines*.

Guidelines for Computing Inverses

1. Write down $y = f(x)$.
2. Solve for x in terms of y.
3. Switch the x's and y's.
4. The result is $y = f^{-1}(x)$.

Example 2.17: Finding the Inverse Function

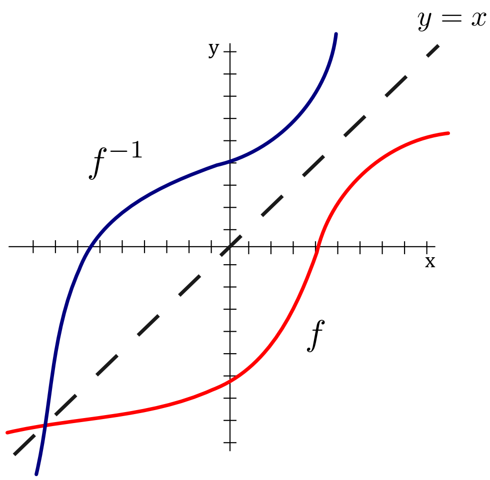
We find the inverse of the function $f(x) = 2x^3 + 1$.

Solution. Starting with $y = 2x^3 + 1$ we solve for x as follows:

$$y - 1 = 2x^3 \qquad \to \qquad \frac{y-1}{2} = x^3 \qquad \to \qquad x = \sqrt[3]{\frac{y-1}{2}}.$$

Therefore, $f^{-1}(x) = \sqrt[3]{\frac{x-1}{2}}$. ♣

This example shows how to find the inverse of a function *algebraically*. But what about finding the inverse of a function *graphically*? Step 3 (switching x and y) gives us a good graphical technique to find the inverse, namely, for each point (a,b) where $f(a) = b$, sketch the point (b,a) for the inverse. More formally, to obtain $f^{-1}(x)$ *reflect* the graph $f(x)$ about the line $y = x$.



Exercises for 2.4

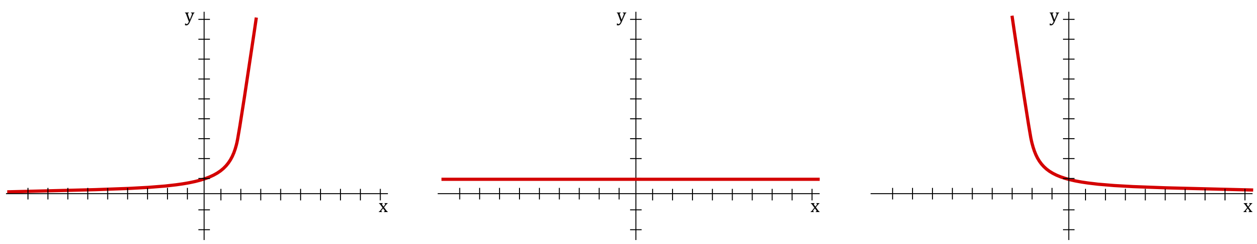
Exercise 2.4.1 *Is the function* $f(x) = |x|$ *one-to-one?*

Exercise 2.4.2 *If $h(x) = e^x + x + 1$, find $h^{-1}(2)$.*

Exercise 2.4.3 *Find a formula for the inverse of the function $f(x) = \dfrac{x+2}{x-2}$.*

2.5 Logarithms

Recall the *three kinds* of exponential functions $f(x) = a^x$ depending on whether $0 < a < 1$, $a = 1$ or $a > 1$:



$f(x) = a^x$ $a > 1$ $\qquad$ $f(x) = 1^x$ $\qquad$ $f(x) = a^x$ $0 < a < 1$

So long as $a \neq 1$, the function $f(x) = a^x$ satisfies the horizontal line test and therefore has an inverse. We call the *inverse of a^x* the **logarithmic function with base a** and denote it by $\log_a$. In particular,

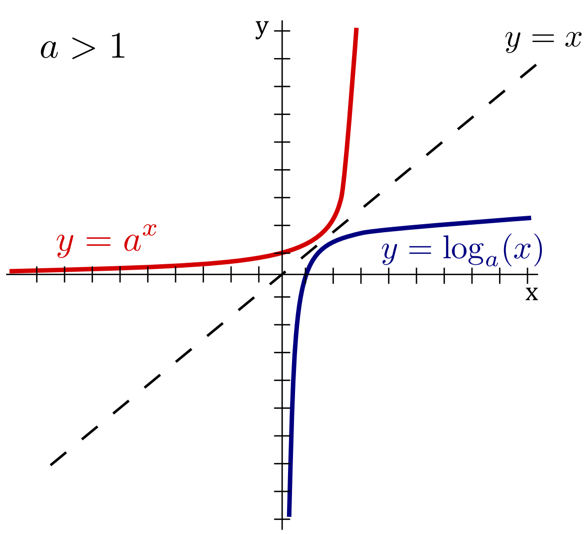
$$\log_a x = y \iff a^y = x.$$

The *cancellation formulas* for logs are:

$$\log_a(a^x) = x, \quad \text{for every } x \in \mathbb{R},$$

$$a^{\log_a(x)} = x, \quad \text{for every } x > 0.$$

Since the function $f(x) = a^x$ for $a \neq 1$ has domain $\mathbb{R}$ and range $(0, \infty)$, the logarithmic function has domain $(0, \infty)$ and range $\mathbb{R}$. For the most part, we only focus on logarithms with a base larger than 1 (i.e., $a > 1$) as these are the most important.



Notice that every logarithm passes through the point $(1,0)$ in the same way that every exponential function passes through the point $(0,1)$.

Some properties of logarithms are as follows.

Logarithm Properties

Let A,B be positive numbers and $b>0$ $(b \neq 1)$ be a base.

- $\log_b(AB) = \log_b A + \log_b B$,
- $\log_b\left(\dfrac{A}{B}\right) = \log_b A - \log_b B$,
- $\log_b(A^n) = n\log_b A$, where n is any real number.

Example 2.18: Compute Lorarithms

To compute $\log_2(24) - \log_2(3)$ *we can do the following:*

$$\log_2(24) - \log_2(3) = \log_2\left(\frac{24}{3}\right) = \log_2(8) = 3,$$

since $2^3 = 8$.

The Natural Logarithm

As mentioned earlier for exponential functions, the number $e \approx 2.71828\ldots$ is the most convenient base to use in Calculus. For this reason we give the logarithm with base e a special name: **the natural logarithm**. We also give it special notation:

$$\log_e x = \ln x.$$

You may pronounce ln as either: "el - en", "lawn", or refer to it as "natural log". The above properties of logarithms also apply to the natural logarithm.

Often we need to turn a logarithm (in a different base) into a natural logarithm. This gives rise to the *change of base formula*.

Change of Base Formula

$$\log_a x = \frac{\ln x}{\ln a}.$$

Example 2.19: Combine Logarithms

Write $\ln A + 2\ln B - \ln C$ *as a single logarithm.*

Solution. Using properties of logarithms, we have,

$$\begin{aligned}\ln A + 2\ln B - \ln C &= \ln A + \ln B^2 - \ln C \\ &= \ln(AB^2) - \ln C \\ &= \ln \frac{AB^2}{C}\end{aligned}$$

♣

Example 2.20: Solve Exponential Equations using Logarithms

If $e^{x+2} = 6e^{2x}$, then solve for x.

Solution. Taking the natural logarithm of both sides and noting the cancellation formulas (along with $\ln e = 1$), we have:

$$\begin{aligned} e^{x+2} &= 6e^{2x} \\ \ln e^{x+2} &= \ln(6e^{2x}) \\ x+2 &= \ln 6 + \ln e^{2x} \\ x+2 &= \ln 6 + 2x \\ x &= 2 - \ln 6 \end{aligned}$$

♣

Example 2.21: Solve Logarithm Equations using Exponentials

If $\ln(2x-1) = 2\ln(x)$, then solve for x.

Solution. "Taking e" of both sides and noting the cancellation formulas, we have:

$$\begin{aligned} e^{\ln(2x-1)} &= e^{2\ln(x)} \\ (2x-1) &= e^{\ln(x^2)} \\ 2x-1 &= x^2 \\ x^2 - 2x + 1 &= 0 \\ (x-1)^2 &= 0 \end{aligned}$$

Therefore, the solution is $x = 1$. ♣

Exercises for 2.5

Exercise 2.5.1 *Expand* $\log_{10}((x+45)^7(x-2))$.

Exercise 2.5.2 *Expand* $\log_2 \dfrac{x^3}{3x-5+(7/x)}$.

Exercise 2.5.3 *Write* $\log_2 3x + 17\log_2(x-2) - 2\log_2(x^2+4x+1)$ *as a single logarithm.*

Exercise 2.5.4 *Solve* $\log_2(1+\sqrt{x}) = 6$ *for x.*

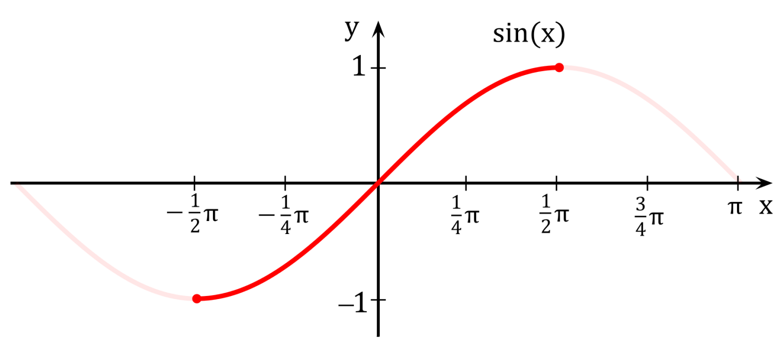
Exercise 2.5.5 *Solve* $2^{x^2} = 8$ *for x.*

Exercise 2.5.6 *Solve* $\log_2(\log_3(x)) = 1$ *for x.*

2.6 Inverse Trigonometric Functions

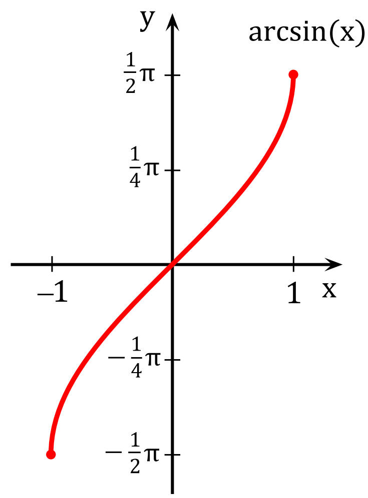
The trigonometric functions frequently arise in problems, and often it is necessary to invert the functions, for example, to find an angle with a specified sine. Of course, there are many angles with the same sine, so the sine function doesn't actually have an inverse that reliably "undoes" the sine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

The sine takes on all values between -1 and 1 exactly once on the interval $[-\pi/2, \pi/2]$.



If we truncate the sine, keeping only the interval $[-\pi/2, \pi/2]$, then this truncated sine has an inverse function. We call this the inverse sine or the arcsine, and write it in one of two common notation: $y =$

arcsin(x), or $y = \sin^{-1}(x)$.



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

Example 2.22: Arcsine of Common Values

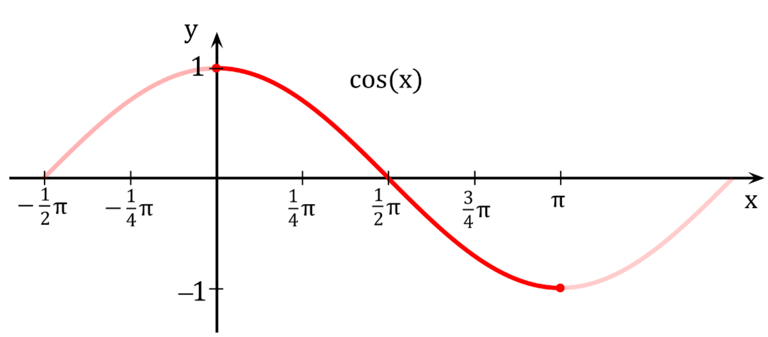
Compute $\sin^{-1}(0)$, $\sin^{-1}(1)$ *and* $\sin^{-1}(-1)$.

Solution. These come directly from the graph of $y = \arcsin x$:

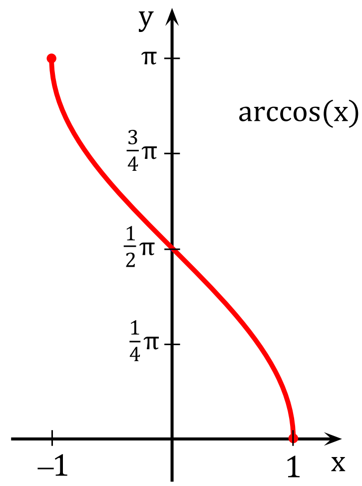
$$\sin^{-1}(0) = 0 \qquad \sin^{-1}(1) = \frac{\pi}{2} \qquad \sin^{-1}(-1) = -\frac{\pi}{2}$$

♣

We can do something similar for the cosine function. As with the sine, we must first truncate the cosine so that it can be inverted, in particular, we use the interval $[0, \pi]$.



Note that the truncated cosine uses a different interval than the truncated sine, so that if $y = \arccos(x)$ we know that $0 \leq y \leq \pi$.



Example 2.23: Arccosine of Common Values

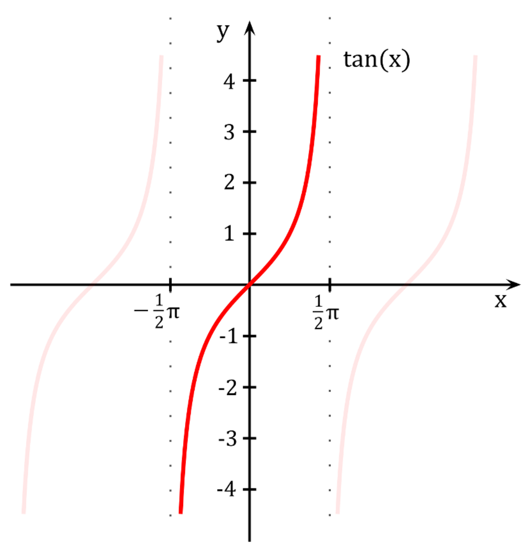
Compute $\cos^{-1}(0)$, $\cos^{-1}(1)$ *and* $\cos^{-1}(-1)$.

Solution. These come directly from the graph of $y = \arccos x$:

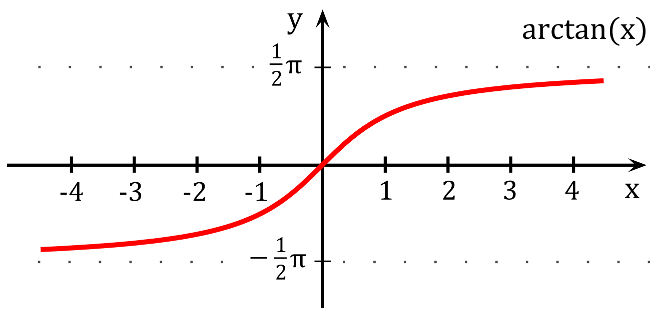
$$\cos^{-1}(0) = \frac{\pi}{2} \qquad \cos^{-1}(1) = 0 \qquad \cos^{-1}(-1) = \pi$$

♣

Finally we look at the tangent; the other trigonometric functions also have "partial inverses" but the sine, cosine and tangent are enough for most purposes. The truncated tangent uses an interval of $(-\pi/2, \pi/2)$.



Reflecting the truncated tangent in the line $y = x$ gives the arctangent function.



Example 2.24: Arctangent of Common Values

Compute $\tan^{-1}(0)$. *What value does* $\tan^{-1} x$ *approach as* x *gets larger and larger? What value does* $\tan^{-1} x$ *approach as* x *gets large (and negative)?*

Solution. These come directly from the graph of $y = \arctan x$. In particular, $\tan^{-1}(0) = 0$. As x gets larger and larger, $\tan^{-1} x$ approaches a value of $\frac{\pi}{2}$, whereas, as x gets large but negative, $\tan^{-1} x$ approaches a value of $-\frac{\pi}{2}$. ♣

The cancellation rules are tricky since we restricted the domains of the trigonometric functions in order to obtain inverse trig functions:

Cancellation Rules

$$\sin(\sin^{-1} x) = x, \quad x \in [-1,1] \qquad \sin^{-1}(\sin x) = x, \quad x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

$$\cos(\cos^{-1} x) = x, \quad x \in [-1,1] \qquad \cos^{-1}(\cos x) = x, \quad x \in [0,\pi]$$

$$\tan(\tan^{-1} x) = x, \quad x \in (-\infty,\infty) \qquad \tan^{-1}(\tan x) = x, \quad x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

Example 2.25: Arcsine of $1/2$

Find $\sin^{-1}(1/2)$.

Solution. Since $\sin^{-1}(x)$ outputs values in $[-\pi/2,\ \pi/2]$, the answer must be in this interval. Let $\theta = \sin^{-1}(1/2)$. We need to compute θ. Take the sine of both sides to get $\sin\theta = \sin(\sin^{-1}(1/2)) = 1/2$ by the cancellation rule. There are many angles θ that work, but we want the one in the interval $[-\pi/2,\ \pi/2]$. Thus, $\theta = \pi/6$ and hence, $\sin^{-1}\left(\frac{1}{2}\right) = \frac{\pi}{6}$. ♣

Example 2.26: Arccosine and the Cancellation Rule

Compute $\cos^{-1}(\cos(0)), \quad \cos^{-1}(\cos(\pi)), \quad \cos^{-1}(\cos(2\pi)), \quad \cos^{-1}(\cos(3\pi)).$

Solution. Since $\cos^{-1}(x)$ outputs values in $[0,\ \pi]$, the answers must be in this interval. The first two we can cancel using the cancellation rules:

$$\cos^{-1}(\cos(0)) = 0 \qquad \text{and} \qquad \cos^{-1}(\cos(\pi)) = \pi.$$

The third one we cannot cancel since $2\pi \notin [0,\pi]$:

$$\cos^{-1}(\cos(2\pi)) \quad \text{is NOT equal to} \quad 2\pi.$$

But we know that cosine is a 2π-periodic function, so $\cos(2\pi) = \cos(0)$:

$$\cos^{-1}(\cos(2\pi)) = \cos^{-1}(\cos(0)) = 0$$

Similarly with the fourth one, we can NOT cancel yet since $3\pi \notin [0,\pi]$. Using $\cos(3\pi) = \cos(3\pi - 2\pi) = \cos(\pi)$:

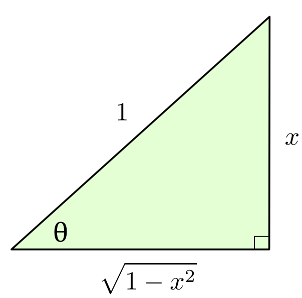
$$\cos^{-1}(\cos(3\pi)) = \cos^{-1}(\cos(\pi)) = \pi.$$

♣

Example 2.27: The Triangle Technique

Rewrite the expression $\cos(\sin^{-1} x)$ *without trig functions. Note that the domain of this function is all* $x \in [-1,1]$.

Solution. Let $\theta = \sin^{-1} x$. We need to compute $\cos\theta$. Taking the sine of both sides gives $\sin\theta = \sin(\sin^{-1}(x)) = x$ by the cancellation rule. We then draw a right triangle using $\sin\theta = x/1$:



If z is the remaining side, then by the Pythagorean Theorem:

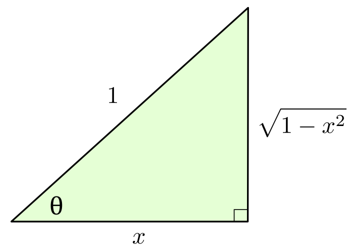
$$z^2 + x^2 = 1 \qquad \to \qquad z^2 = 1 - x^2 \qquad \to \qquad z = \pm\sqrt{1-x^2}$$

and hence $z = +\sqrt{1-x^2}$ since $\theta \in [-\pi/2,\ \pi/2]$. Thus, $\cos\theta = \sqrt{1-x^2}$ by SOH CAH TOA, so, $\cos(\sin^{-1} x) = \sqrt{1-x^2}$. ♣

Example 2.28: The Triangle Technique 2

For $x \in (0,1)$, *rewrite the expression* $\sin(2\cos^{-1} x)$. *Compute* $\sin(2\cos^{-1}(1/2))$.

Solution. Let $\theta = \cos^{-1} x$ so that $\cos\theta = x$. The question now asks for us to compute $\sin(2\theta)$. We then draw a right triangle using $\cos\theta = x/1$:



To find $\sin(2\theta)$ we use the double angle formula $\sin(2\theta) = 2\sin\theta\cos\theta$. But $\sin\theta = \sqrt{1-x^2}$, for $\theta \in [0,\pi]$, and $\cos\theta = x$. Therefore, $\sin(2\cos^{-1} x) = 2x\sqrt{1-x^2}$. When $x = 1/2$ we have $\sin(2\cos^{-1}(1/2)) = \dfrac{\sqrt{3}}{2}$. ♣

Exercises for 2.6

Exercise 2.6.1 *Compute the following:*

(a) $\sin^{-1}(\sqrt{3}/2)$

(b) $\cos^{-1}(-\sqrt{2}/2)$

Exercise 2.6.2 *Compute the following:*

(a) $\sin^{-1}(\sin(\pi/4))$

(b) $\sin^{-1}(\sin(17\pi/3))$

(c) $\cos\left(\cos^{-1}(1/3)\right)$

(d) $\tan\left(\cos^{-1}(-4/5)\right)$

Exercise 2.6.3 *Rewrite the expression* $\tan\left(\cos^{-1} x\right)$ *without trigonometric functions. What is the domain of this function?*

2.7 Hyperbolic Functions

The hyperbolic functions appear with some frequency in applications, and are quite similar in many respects to the trigonometric functions. This is a bit surprising given our initial definitions.

Definition 2.29: Hyperbolic Sine and Cosine

The **hyperbolic cosine** *is the function*

$$\cosh x = \frac{e^x + e^{-x}}{2},$$

and the **hyperbolic sine** *is the function*

$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

Notice that cosh is even (that is, $\cosh(-x) = \cosh(x)$) while sinh is odd ($\sinh(-x) = -\sinh(x)$), and $\cosh x + \sinh x = e^x$. Also, for all x, $\cosh x > 0$, while $\sinh x = 0$ if and only if $e^x - e^{-x} = 0$, which is true precisely when $x = 0$.

Theorem 2.30: Range of Hyperbolic Cosine

The range of $\cosh x$ *is* $[1, \infty)$.

Proof. Let $y = \cosh x$. We solve for x:

$$\begin{aligned} y &= \frac{e^x + e^{-x}}{2} \\ 2y &= e^x + e^{-x} \\ 2ye^x &= e^{2x} + 1 \\ 0 &= e^{2x} - 2ye^x + 1 \\ e^x &= \frac{2y \pm \sqrt{4y^2 - 4}}{2} \\ e^x &= y \pm \sqrt{y^2 - 1} \end{aligned}$$

From the last equation, we see $y^2 \geq 1$, and since $y \geq 0$, it follows that $y \geq 1$.

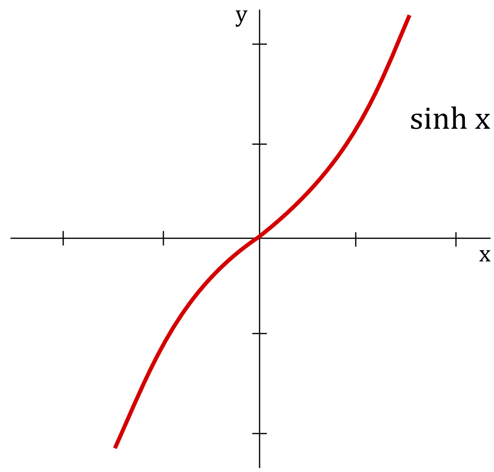
Now suppose $y \geq 1$, so $y \pm \sqrt{y^2 - 1} > 0$. Then $x = \ln(y \pm \sqrt{y^2 - 1})$ is a real number, and $y = \cosh x$, so y is in the range of $\cosh(x)$. ♣

Definition 2.31: Hyperbolic Functions

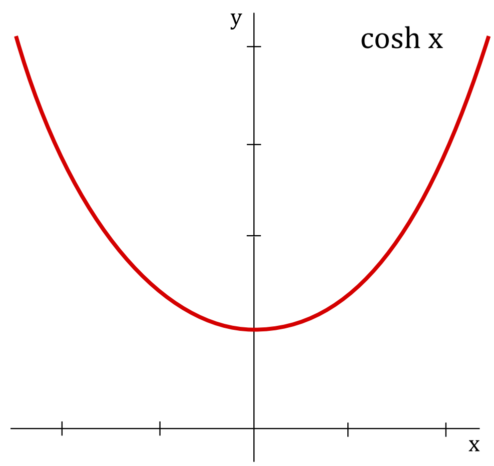
We can also define hyperbolic functions for the other trigonometric functions as you would expect:

$$\tanh x = \frac{\sinh x}{\cosh x} \qquad \operatorname{csch} x = \frac{1}{\sinh x} \qquad \operatorname{sech} x = \frac{1}{\cosh x} \qquad \coth x = \frac{1}{\tanh x}$$

The graph of $\sinh x$ is shown below:



The graph of $\cosh x$ is shown below:



Example 2.32: Computing Hyperbolic Tangent

Compute $\tanh(\ln 2)$.

Solution. This uses the definitions of the hyperbolic functions.

$$\tanh(\ln 2) = \frac{\sinh(\ln 2)}{\cosh(\ln 2)} = \frac{\dfrac{e^{\ln 2} - e^{-\ln 2}}{2}}{\dfrac{e^{\ln 2} + e^{-\ln 2}}{2}} = \frac{\dfrac{2 - (1/2)}{2}}{\dfrac{2 + (1/2)}{2}} = \frac{2 - (1/2)}{2 + (1/2)} = \frac{3}{5}$$

♣

Certainly the hyperbolic functions do not closely resemble the trigonometric functions graphically. But they do have analogous properties, beginning with the following identity.

Theorem 2.33: Hyperbolic Identity

For all x in $\mathbb{R}$, $\cosh^2 x - \sinh^2 x = 1$.

Proof. The proof is a straightforward computation:

$$\cosh^2 x - \sinh^2 x = \frac{(e^x + e^{-x})^2}{4} - \frac{(e^x - e^{-x})^2}{4} = \frac{e^{2x} + 2 + e^{-2x} - e^{2x} + 2 - e^{-2x}}{4} = \frac{4}{4} = 1.$$

♣

This immediately gives two additional identities:

$$1 - \tanh^2 x = \operatorname{sech}^2 x \qquad \text{and} \qquad \coth^2 x - 1 = \operatorname{csch}^2 x.$$

The identity of the theorem also helps to provide a geometric motivation. Recall that the graph of $x^2 - y^2 = 1$ is a hyperbola with asymptotes $x = \pm y$ whose x-intercepts are ± 1. If (x, y) is a point on the right half of the hyperbola, and if we let $x = \cosh t$, then $y = \pm\sqrt{x^2 - 1} = \pm\sqrt{\cosh^2 x - 1} = \pm \sinh t$. So for some suitable t, $\cosh t$ and $\sinh t$ are the coordinates of a typical point on the hyperbola. In fact, it turns out that t is twice the area shown in the first graph of figure 2.4. Even this is analogous to trigonometry; $\cos t$ and $\sin t$ are the coordinates of a typical point on the unit circle, and t is twice the area shown in the second graph of Figure 2.4.

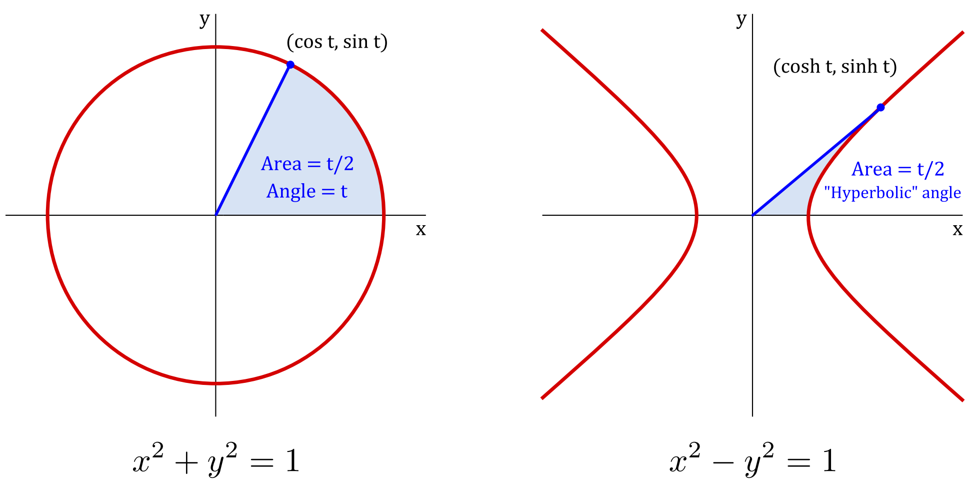


Figure 2.4: Geometric definitions. Here, t is twice the shaded area in each figure.

Since $\cosh x > 0$, $\sinh x$ is increasing and hence one-to-one, so $\sinh x$ has an inverse, $\operatorname{arcsinh} x$. Also, $\sinh x > 0$ when $x > 0$, so $\cosh x$ is injective on $[0, \infty)$ and has a (partial) inverse, $\operatorname{arccosh} x$. The other hyperbolic functions have inverses as well, though $\operatorname{arcsech} x$ is only a partial inverse.

Exercises for 2.7

Exercise 2.7.1 *Show that the range of $\sinh x$ is all real numbers. (Hint: show that if $y = \sinh x$ then $x = \ln(y + \sqrt{y^2 + 1})$.)*

Exercise 2.7.2 *Show that the range of* $\tanh x$ *is* $(-1,1)$. *What are the ranges of* coth, sech, *and* csch*? (Use the fact that they are reciprocal functions.)*

Exercise 2.7.3 *Prove that for every* $x, y \in \mathbb{R}$, $\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y$. *Obtain a similar identity for* $\sinh(x-y)$.

Exercise 2.7.4 *Prove that for every* $x, y \in \mathbb{R}$, $\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y$. *Obtain a similar identity for* $\cosh(x-y)$.

Exercise 2.7.5 *Show that* $\sinh(2x) = 2\sinh x \cosh x$ *and* $\cosh(2x) = \cosh^2 x + \sinh^2 x$ *for every* x. *Conclude also that* $(\cosh(2x)-1)/2 = \sinh^2 x$.

Exercise 2.7.6 *What are the domains of the six inverse hyperbolic functions?*

Exercise 2.7.7 *Sketch the graphs of all six inverse hyperbolic functions.*

2.8 Additional Exercises

Exercise 2.8.1 *If* $f(x) = \dfrac{1}{x-1}$, *then which of the following is equal to* $f\left(\dfrac{1}{x}\right)$?

(a) $f(x)$

(b) $-f(x)$

(c) $xf(x)$

(d) $-xf(x)$

(e) $\dfrac{f(x)}{x}$

(f) $-\dfrac{f(x)}{x}$

Exercise 2.8.2 *If* $f(x) = \dfrac{x}{x+3}$, *then find and simplify* $\dfrac{f(x)-f(2)}{x-2}$.

Exercise 2.8.3 *If* $f(x) = x^2$, *then find and simplify* $\dfrac{f(3+h)-f(3)}{h}$.

Exercise 2.8.4 *What is the domain of*

(a) $f(x) = \dfrac{\sqrt{x-2}}{x^2-9}$?

(b) $g(x) = \dfrac{\sqrt[3]{x-2}}{x^2-9}$?

Exercise 2.8.5 *Suppose that $f(x) = x^3$ and $g(x) = x$. What is the domain of $\dfrac{f}{g}$?*

Exercise 2.8.6 *Suppose that $f(x) = 3x-4$. Find a function g such that $(g \circ f)(x) = 5x+2$.*

Exercise 2.8.7 *Which of the following functions is one-to-one?*

(a) $f(x) = x^2+4x+3$

(b) $g(x) = |x|+2$

(c) $h(x) = \sqrt[3]{x+1}$

(d) $F(x) = \cos x,\ -\pi \leq x \leq \pi$

(e) $G(x) = e^x + e^{-x}$

Exercise 2.8.8 *What is the inverse of $f(x) = \ln\left(\dfrac{e^x}{e^x-1}\right)$? What is the domain of f^{-1}?*

Exercise 2.8.9 *Solve the following equations.*

(a) $e^{2-x} = 3$

(b) $e^{x^2} = e^{4x-3}$

(c) $\ln(1+\sqrt{x}) = 2$

(d) $\ln(x^2-3) = \ln 2 + \ln x$

Exercise 2.8.10 *Find the exact value of* $\sin^{-1}\left(-\sqrt{2}/2\right) - \cos^{-1}\left(-\sqrt{2}/2\right)$.

Exercise 2.8.11 *Find* $\sin^{-1}(\sin(23\pi/5))$.

Exercise 2.8.12 *It can be proved that $f(x) = x^3+x+e^{x-1}$ is one-to-one. What is the value of $f^{-1}(3)$?*

Exercise 2.8.13 *Sketch the graph of* $f(x) = \begin{cases} -x & \text{if } x \leq 0 \\ \tan^{-1} x & \text{if } x > 0 \end{cases}$

3. Limits

3.1 The Limit

The value a function f approaches as its input x approaches some value is said to be the limit of f. Limits are essential to the study of calculus and, as we will see, are used in defining continuity, derivatives, and integrals.

Consider the function

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

Notice that $x = 1$ does not belong to the domain of $f(x)$. Regardless, we would like to know how $f(x)$ behaves close to the point $x = 1$. We start with a table of values:

x	$f(x)$
0.5	1.5
0.9	1.9
0.99	1.99
1.01	2.01
1.1	2.1
1.5	2.5

It appears that for values of x close to 1 we have that $f(x)$ is close to 2. In fact, we can make the values of $f(x)$ as close to 2 as we like by taking x sufficiently close to 1. We express this by saying *the limit of the function $f(x)$ as x approaches 1 is equal to 2* and use the notation:

$$\lim_{x \to 1} f(x) = 2.$$

Definition 3.1: Limit (Useable Definition)

In general, we will write

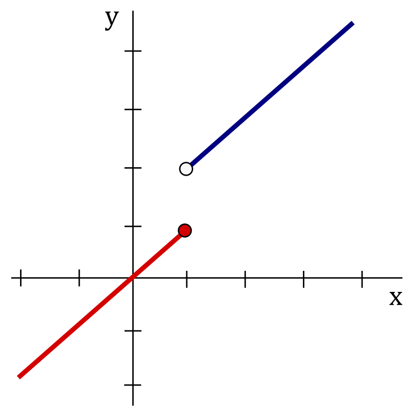
$$\lim_{x \to a} f(x) = L,$$

if we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a (on either side of a) but not equal to a.

We read the expression $\lim_{x \to a} f(x) = L$ as "*the limit of $f(x)$ as x approaches a is equal to L*". When evaluating a limit, you are essentially answering the following question: What number does the function *approach* while x gets closer and closer to a (but *not equal* to a)? The phrase *but not equal to a* in the definition of a limit means that when finding the limit of $f(x)$ as x approaches a we never actually consider $x = a$. In fact, as we just saw in the example above, a may not even belong to the domain of f. All that matters for limits is what happens to f close to a, not necessarily what happens to f at a.

One-sided limits

Consider the following piecewise defined function:



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

Observe from the graph that as x gets closer and closer to 1 from the *left*, then $f(x)$ approaches $+1$. Similarly, as x gets closer and closer 1 from the *right*, then $f(x)$ approaches $+2$. We use the following notation to indicate this:

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

The symbol $x \to 1^-$ means that we only consider values of x sufficiently close to 1 which are less than 1. Similarly, the symbol $x \to 1^+$ means that we only consider values of x sufficiently close to 1 which are greater than 1.

Definition 3.2: Left and Right-Hand Limit (Useable Definition)

In general, we will write

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

if we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a and x less than a. This is called the **left-hand limit** *of $f(x)$ as x approaches a. Similarly, we write*

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

if we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a and x greater than a. This is called the **right-hand limit** *of $f(x)$ as x approaches a.*

We note the following fact:

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

Or more concisely:

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

A consequence of this fact is that if the one-sided limits are *different*, then the two-sided limit $\lim_{x\to a} f(x)$ does not exist, often denoted as: (DNE).

Exercises for Section 3.1

Exercise 3.1.1 *Use a calculator to estimate* $\lim_{x\to 0} \dfrac{\sin x}{x}$, *where x is in radians.*

Exercise 3.1.2 *Use a calculator to estimate* $\lim_{x\to 0} \dfrac{\tan(3x)}{\tan(5x)}$, *where x is in radians.*

Exercise 3.1.3 *Use a calculator to estimate* $\lim_{x\to 1^+} \dfrac{|x-1|}{1-x^2}$ *and* $\lim_{x\to 1^-} \dfrac{|x-1|}{1-x^2}$.

3.2 Precise Definition of a Limit

The definiton given for a limit previously is more of a working definition. In this section we pursue the actual, official definition of a limit.

Definition 3.3: Precise Definition of Limit

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

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

Make sure you are not confused by the names of important quantities. The generic definition talks about $f(x)$, but the function and the variable might have other names. The x was the variable of the original function; when we were trying to compute a slope or a velocity, x was essentially a fixed quantity, telling us at what point we wanted the slope. In the velocity problem, it was literally a fixed quantity, as we focused on the time $t = 2$. The quantity a of the definition in all the examples was zero: we were always interested in what happened as Δx became very close to zero.

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

Example 3.4: Epsilon Delta

Let's show carefully that $\lim_{x\to 2} x+4=6$.

Solution. This is not something we "need" to prove, since it is "obviously" true. But if we couldn't prove it using our official definition there would be something very wrong with the definition.

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

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

Example 3.5: Epsilon Delta

It seems clear that $\lim_{x\to 2} x^2=4$. *Let's try to prove it.*

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

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

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

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

Theorem 3.6: Limit Product

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

Proof. We must use the Precise Definition of a Limit to prove the Produce Law for Limits. So given any ε we need to find a δ so that $0 < |x-a| < \delta$ implies $|f(x)g(x) - LM| < \varepsilon$. What do we have to work with? We know that we can make $f(x)$ close to L and $g(x)$ close to M, and we have to somehow connect these facts to make $f(x)g(x)$ close to LM.

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

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

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

Suppose $\varepsilon > 0$. Since $\lim_{x\to a} f(x) = L$, there is a value δ_1 such that $0 < |x-a| < \delta_1$ implies $|f(x) - L| < \dfrac{\varepsilon}{2(1+|M|)}$. This means that $0 < |x-a| < \delta_1$ implies $|f(x) - L||M| < |f(x) - L|(1+|M|) < \varepsilon/2$.

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

$$\begin{aligned} |f(x)g(x) - LM| &\le |f(x)||g(x) - M| + |f(x) - L||M| \\ &< N\frac{\varepsilon}{2N} + \frac{\varepsilon}{2(1+|M|)}(1+|M|) \\ &= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

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

The concept of a **one-sided limit** can also be made precise.

Definition 3.7: One-sided Limit

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

Exercises for Section 3.2

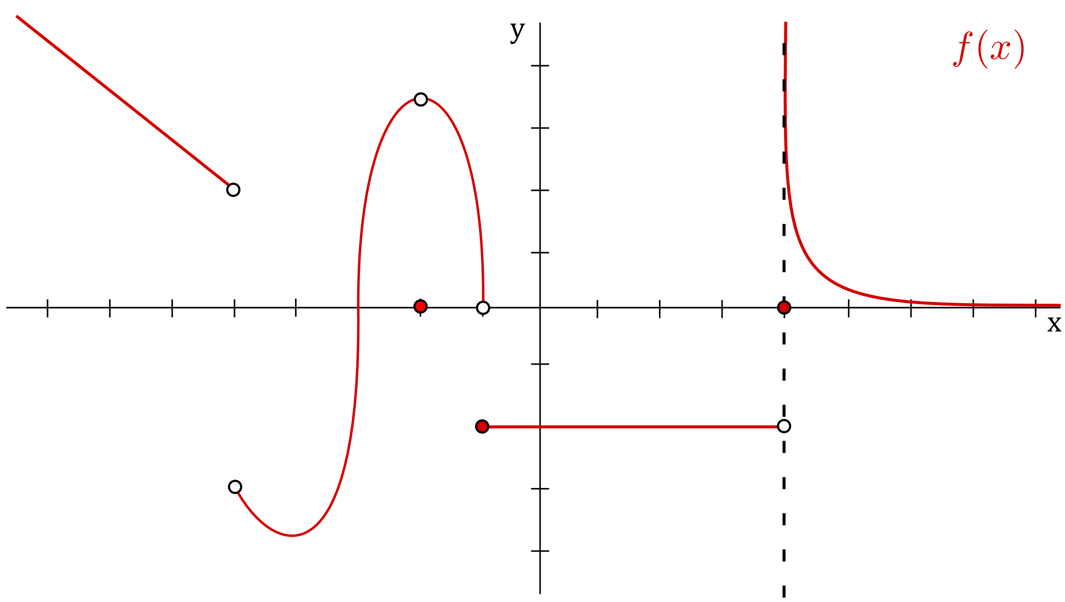
Exercise 3.2.1 *Give an ε–δ proof of the fact that* $\lim_{x\to 4}(2x-5) = 3$.

Exercise 3.2.2 *Let ε be a small positive real number. How close to 2 must we hold x in order to be sure that $3x+1$ lies within ε units of 7?*

3.3 Computing Limits: Graphically

In this section we look at an example to illustrate the concept of a limit *graphically*.

The graph of a function $f(x)$ is shown below. We will analyze the behaviour of $f(x)$ around $x = -5$, $x = -2$, $x = -1$ and $x = 0$, and $x = 4$.



Observe that $f(x)$ is indeed a function (it passes the vertical line test). We now analyze the function at each point separately.

$x = -5$: Observe that at $x = -5$ there is no closed circle, thus $f(-5)$ is undefined. From the graph we see that as x gets closer and closer to -5 from the left, then $f(x)$ approaches 2, so

$$\lim_{x \to -5^-} f(x) = 2.$$

Similarly, as x gets closer and closer -5 from the right, then $f(x)$ approaches -3, so

$$\lim_{x \to -5^+} f(x) = -3.$$

As the right-hand limit and left-hand limit are not equal at -5, we know that

$$\lim_{x \to -5} f(x) \quad \text{does not exist.}$$

$x = -2$: Observe that at $x = -2$ there is a closed circle at 0, thus $f(-2) = 0$. From the graph we see that as x gets closer and closer to -2 from the left, then $f(x)$ approaches 3.5, so

$$\lim_{x \to -2^-} f(x) = 3.5.$$

Similarly, as x gets closer and closer -2 from the right, then $f(x)$ again approaches 3.5, so

$$\lim_{x \to -2^+} f(x) = 3.5.$$

As the right-hand limit and left-hand limit are both equal to 3.5, we know that

$$\lim_{x \to -2} f(x) = 3.5.$$

Do not be concerned that the limit does not equal 0. This is a discontinuity, which is completely valid, and will be discussed in a later section.

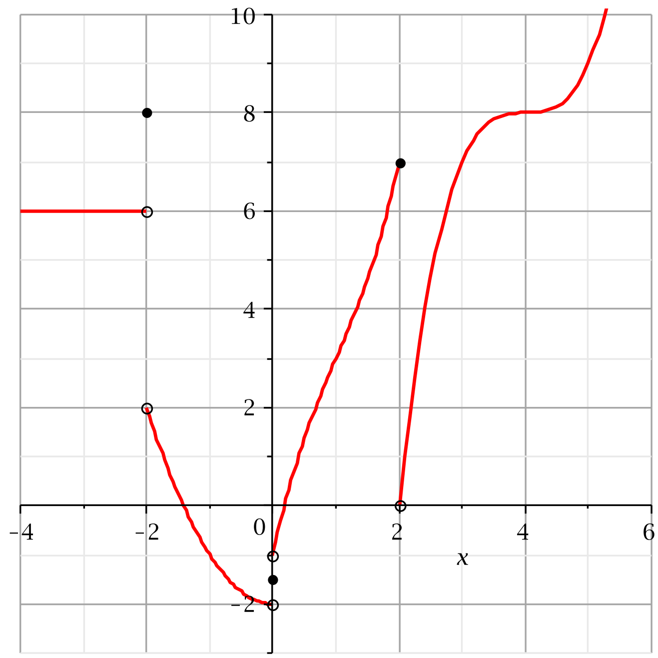
We leave it to the reader to analyze the behaviour of $f(x)$ for x close to -1 and 0.

Summarizing, we have:

$f(-5)$ is undefined	$f(-2) = 0$	$f(-1) = -2$	$f(0) = -2$
$\lim_{x \to -5^-} f(x) = 2$	$\lim_{x \to -2^-} f(x) = 3.5$	$\lim_{x \to -1^-} f(x) = 0$	$\lim_{x \to 0^-} f(x) = -2$
$\lim_{x \to -5^+} f(x) = -3$	$\lim_{x \to -2^+} f(x) = 3.5$	$\lim_{x \to -1^+} f(x) = -2$	$\lim_{x \to 0^+} f(x) = -2$
$\lim_{x \to -5} f(x) = \text{DNE}$	$\lim_{x \to -2} f(x) = 3.5$	$\lim_{x \to -1} f(x) = \text{DNE}$	$\lim_{x \to 0} f(x) = -2$

Exercises for Section 3.3

Exercise 3.3.1 *Evaluate the expressions by reference to this graph:*



(a) $\lim_{x\to 4} f(x)$

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

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

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

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

(f) $f(-2)$

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

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

(i) $\lim_{x\to 0} f(x+1)$

(j) $f(0)$

(k) $\lim_{x\to 1^-} f(x-4)$

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

3.4 Computing Limits: Algebraically

Properties of limits

We begin by deriving a handful of theorems to give us the tools to compute many limits without explicitly working with the precise definition of a limit.

Theorem 3.8: Limit Properties

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

- $\lim_{x\to a} kf(x) = k\lim_{x\to a} f(x) = kL$
- $\lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = L+M$
- $\lim_{x\to a}(f(x)-g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = L-M$
- $\lim_{x\to a}(f(x)g(x)) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = LM$
- $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \dfrac{L}{M}$, *if* M *is not 0*

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

Example 3.9: Limit Properties

Compute $\lim_{x\to 1} \dfrac{x^2-3x+5}{x-2}$.

Solution. If we apply the theorem in all its gory detail, we get

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

♣

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

We're primarily interested in limits that aren't so easy, namely, limits in which a denominator approaches zero. There are a handful of algebraic tricks that work on many of these limits.

Example 3.10: Zero Denominator

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

Solution. We can't simply plug in $x = 1$ because that makes the denominator zero. However:

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

♣

The technique used to solve the previous example can be referred to as *factor and cancel*. Its validity comes from the fact that we are allowed to cancel $x-1$ from the numerator and denominator. Remember in Calculus that we have to make sure we don't cancel zeros, so we require $x-1 \neq 0$ in order to cancel it. But looking back at the definition of a limit using $x \to 1$, the key point for this example is that we are taking values of x close to 1 but *not* equal to 1. This is exactly what we wanted ($x \neq 1$) in order to cancel this common factor.

While Theorem 3.8 is very helpful, we need a bit more to work easily with limits. Since the theorem applies when some limits are already known, we need to know the behavior of some functions that cannot themselves be constructed from the simple arithmetic operations of the theorem, such as $\sqrt{x}$. Also, there is one other extraordinarily useful way to put functions together: composition. If $f(x)$ and $g(x)$ are functions, we can form two functions by composition: $f(g(x))$ and $g(f(x))$. For example, if $f(x) = \sqrt{x}$ and $g(x) = x^2+5$, then $f(g(x)) = \sqrt{x^2+5}$ and $g(f(x)) = (\sqrt{x})^2+5 = x+5$. Here is a companion to Theorem 3.8 for composition:

Theorem 3.11: Limit of Composition

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

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

Note the special form of the condition on f: it is not enough to know that $\lim_{x\to L} f(x) = M$, though it is a bit tricky to see why. We have included an example in the exercise section to illustrate this tricky point for those who are interested. Many of the most familiar functions do have this property, and this theorem can therefore be applied. For example:

Theorem 3.12: Continuity of Roots

Suppose that n is a positive integer. Then

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

provided that a is positive if n is even.

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

Another of the most common algebraic tricks is called *rationalization*. Rationalizing makes use of the difference of squares formula $(a-b)(a+b) = a^2 - b^2$. Here is an example.

Example 3.13: Rationalizing

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

Solution.

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

At the very last step we have used Theorems 3.11 and 3.12. ♣

Example 3.14: Left and Right Limit

Evaluate $\displaystyle\lim_{x\to 0^+} \frac{x}{|x|}$.

Solution. The function $f(x) = x/|x|$ is undefined at 0; when $x > 0$, $|x| = x$ and so $f(x) = 1$; when $x < 0$, $|x| = -x$ and $f(x) = -1$. Thus

$$\lim_{x\to 0^-} \frac{x}{|x|} = \lim_{x\to 0^-} -1 = -1$$

while

$$\lim_{x\to 0^+} \frac{x}{|x|} = \lim_{x\to 0^+} 1 = 1.$$

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

Exercises for 3.4

Exercise 3.4.1 *Compute the limits. If a limit does not exist, explain why.*

(a) $\displaystyle\lim_{x\to 3} \frac{x^2+x-12}{x-3}$

(b) $\displaystyle\lim_{x\to 1} \frac{x^2+x-12}{x-3}$

(c) $\displaystyle\lim_{x\to -4} \frac{x^2+x-12}{x-3}$

(d) $\displaystyle\lim_{x\to 2} \frac{x^2+x-12}{x-2}$

(e) $\displaystyle\lim_{x\to 1} \frac{\sqrt{x+8}-3}{x-1}$

(f) $\displaystyle\lim_{x\to 0^+} \sqrt{\frac{1}{x}+2}-\sqrt{\frac{1}{x}}$

(g) $\displaystyle\lim_{x\to 2} 3$

(h) $\displaystyle\lim_{x\to 4} 3x^3-5x$

(i) $\displaystyle\lim_{x\to 0} \frac{4x-5x^2}{x-1}$

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

(k) $\displaystyle\lim_{x\to 0^+} \frac{\sqrt{2-x^2}}{x}$

(l) $\displaystyle\lim_{x\to 0^+} \frac{\sqrt{2-x^2}}{x+1}$

(m) $\displaystyle\lim_{x\to a} \frac{x^3-a^3}{x-a}$

(n) $\displaystyle\lim_{x\to 2} (x^2+4)^3$

Exercise 3.4.2 *Let* $f(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$ *and* $g(x) = 0$. *What are the values of* $L = \lim_{x\to 0} g(x)$ *and* $M = \lim_{x\to L} f(x)$*? Is it true that* $\lim_{x\to 0} f(g(x)) = M$*? What are some noteworthy differences between this example and Theorem 3.11?*

3.5 Infinite Limits and Limits at Infinity

We occasionally want to know what happens to some quantity when a variable gets very large or "goes to infinity".

Example 3.15: Limit at Infinity

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

As with ordinary limits, this concept of "limit at infinity" can be made precise. Roughly, we want $\lim\limits_{x\to\infty} f(x) = L$ to mean that we can make $f(x)$ as close as we want to L by making x large enough.

Definition 3.16: Limit at Infinity (Formal Definition)

If f is a function, we say that $\lim_{x\to\infty} f(x) = L$ if for every $\varepsilon > 0$ there is an $N > 0$ so that whenever $x > N$, $|f(x) - L| < \varepsilon$. We may similarly define $\lim_{x\to-\infty} f(x) = L$.

We include this definition for completeness, but we will not explore it in detail. Suffice it to say that such limits behave in much the same way that ordinary limits do; in particular there is a direct analog of Theorem 3.8.

Example 3.17: Limit at Infinity

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

Solution. As x goes to infinity both the numerator and denominator go to infinity. We divide the numerator and denominator by x^2:

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

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

In the previous example, we *divided by the highest power of x that occurs in the denominator* in order to evaluate the limit. We illustrate another technique similar to this.

Example 3.18: Limit at Infinity

Compute the following limit:

$$\lim_{x\to\infty} \frac{2x^2 + 3}{5x^2 + x}.$$

Solution. As x becomes large, both the numerator and denominator become large, so it isn't clear what happens to their ratio. The highest power of x in the denominator is x^2, therefore we will divide every term in both the numerator and denominator by x^2 as follows:

$$\lim_{x\to\infty} \frac{2x^2 + 3}{5x^2 + x} = \lim_{x\to\infty} \frac{2 + 3/x^2}{5 + 1/x}.$$

Most of the limit rules from last lecture also apply to infinite limits, so we can write this as:

$$= \frac{\lim\limits_{x\to\infty} 2 + 3 \lim\limits_{x\to\infty} \dfrac{1}{x^2}}{\lim\limits_{x\to\infty} 5 + \lim\limits_{x\to\infty} \dfrac{1}{x}} = \frac{2 + 3(0)}{5 + 0} = \frac{2}{5}.$$

Note that we used the theorem above to get that $\lim_{x\to\infty} \dfrac{1}{x} = 0$ and $\lim_{x\to\infty} \dfrac{1}{x^2} = 0$.

A shortcut technique is to analyze only the *leading terms* of the numerator and denominator. A leading term is a term that has the highest power of x. If there are multiple terms with the same exponent, you must include all of them.
Top: The leading term is $2x^2$.
Bottom: The leading term is $5x^2$.
Now only looking at leading terms and ignoring the other terms we get:

$$\lim_{x\to\infty} \frac{2x^2+3}{5x^2+x} = \lim_{x\to\infty} \frac{2x^2}{5x^2} = \frac{2}{5}.$$

♣

We next look at limits whose value is infinity (or minus infinity).

Definition 3.19: Infinite Limit (Useable Definition)

In general, we will write

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

if we can make the value of $f(x)$ arbitrarily large by taking x to be sufficiently close to a (on either side of a) but not equal to a. Similarly, we will write

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

if we can make the value of $f(x)$ arbitrarily large and negative by taking x to be sufficiently close to a (on either side of a) but not equal to a.

This definition can be modified for one-sided limits as well as limits with $x \to a$ replaced by $x \to \infty$ or $x \to -\infty$.

Example 3.20: Limit at Infinity

Compute the following limit: $\lim\limits_{x\to\infty} (x^3 - x)$.

Solution. One might be tempted to write:

$$\lim_{x\to\infty} x^3 - \lim_{x\to\infty} x = \infty - \infty,$$

however, we do not know what $\infty - \infty$ is, as ∞ is not a real number and so cannot be treated like one. We instead write:

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

As x becomes arbitrarily large, then both x and $x^2 - 1$ become arbitrarily large, and hence their product $x(x^2 - 1)$ will also become arbitrarily large. Thus we see that

$$\lim_{x\to\infty} (x^3 - x) = \infty.$$

♣

Example 3.21: Limit at Infinity and Basic Functions

We can easily evaluate the following limits by observation:

1. $\lim_{x\to\infty} \frac{6}{\sqrt{x^3}} = 0$
2. $\lim_{x\to-\infty} x - x^2 = -\infty$
3. $\lim_{x\to\infty} x^3 + x = \infty$
4. $\lim_{x\to\infty} \cos(x) = DNE$
5. $\lim_{x\to\infty} e^x = \infty$
6. $\lim_{x\to-\infty} e^x = 0$
7. $\lim_{x\to 0^+} \ln x = -\infty$
8. $\lim_{x\to 0} \cos(1/x) = DNE$

Often, the shorthand notation $\frac{1}{0^+} = +\infty$ and $\frac{1}{0^-} = -\infty$ is used to represent the following two limits respectively:

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

Using the above convention we can compute the following limits.

Example 3.22: Limit at Infinity and Basic Functions

Compute $\lim_{x\to 0^+} e^{1/x}$, $\lim_{x\to 0^-} e^{1/x}$ *and* $\lim_{x\to 0} e^{1/x}$.

Solution. We have:

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

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

Thus, as left-hand limit $\neq$ right-hand limit,

$$\lim_{x\to 0} e^{\frac{1}{x}} = \text{DNE}.$$

♣

3.5.1. Vertical Asymptotes

The line $x = a$ is called a **vertical asymptote** of $f(x)$ if *at least one* of the following is true:

$$\lim_{x\to a} f(x) = \infty \qquad \lim_{x\to a^-} f(x) = \infty \qquad \lim_{x\to a^+} f(x) = \infty$$

$$\lim_{x\to a} f(x) = -\infty \qquad \lim_{x\to a^-} f(x) = -\infty \qquad \lim_{x\to a^+} f(x) = -\infty$$

Example 3.23: Vertical Asymptotes

Find the vertical asymptotes of $f(x) = \dfrac{2x}{x-4}$.

Solution. In the definition of vertical asymptotes we need a certain limit to be $\pm\infty$. Candidates would be to consider values not in the domain of $f(x)$, such as $a = 4$. As x approaches 4 but is larger than 4 then $x-4$ is a small positive number and $2x$ is close to 8, so the quotient $2x/(x-4)$ is a large positive number. Thus we see that

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

Thus, at least one of the conditions in the definition above is satisfied. Therefore $x = 4$ is a vertical asymptote. ♣

3.5.2. Horizontal Asymptotes

The line $y = L$ is a **horizontal asymptote** of $f(x)$ if either

$$\lim_{x\to\infty} f(x) = L \qquad \text{or} \qquad \lim_{x\to-\infty} f(x) = L.$$

Example 3.24: Horizontal Asymptotes

Find the horizontal asymptotes of $f(x) = \dfrac{|x|}{x}$.

Solution. We must compute two infinite limits. First,

$$\lim_{x\to\infty} \frac{|x|}{x}.$$

Notice that for x arbitrarily large that $x > 0$, so that $|x| = x$. In particular, for x in the interval $(0,\infty)$ we have

$$\lim_{x\to\infty} \frac{|x|}{x} = \lim_{x\to\infty} \frac{x}{x} = 1.$$

Second, we must compute

$$\lim_{x\to-\infty} \frac{|x|}{x}.$$

Notice that for x arbitrarily large negative that $x < 0$, so that $|x| = -x$. In particular, for x in the interval $(-\infty,0)$ we have

$$\lim_{x\to-\infty} \frac{|x|}{x} = \lim_{x\to-\infty} \frac{-x}{x} = -1.$$

Therefore there are two horizontal asymptotes, namely, $y = 1$ and $y = -1$. ♣

3.5.3. Slant Asymptotes

Some functions may have slant (or *oblique*) asymptotes, which are neither vertical nor horizontal. If

$$\lim_{x\to\infty}[f(x)-(mx+b)]=0$$

then the straight line $y=mx+b$ is a **slant asymptote** to $f(x)$. Visually, the vertical distance between $f(x)$ and $y=mx+b$ is decreasing towards 0 and the curves do not intersect or cross at any point as x approaches infinity. Similarly when $x\to-\infty$.

Example 3.25: Slant Asymptote in a Rational Function

Find the slant asymptotes of $f(x)=\dfrac{-3x^2+4}{x-1}$.

Solution. Note that this function has no horizontal asymptotes since $f(x)\to-\infty$ as $x\to\infty$ and $f(x)\to\infty$ as $x\to-\infty$.

In rational functions, slant asymptotes occur when the degree in the numerator is one greater than in the denominator. We use long division to rearrange the function:

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

The part we're interested in is the resulting polynomial $-3x-3$. This is the line $y=mx+b$ we were seeking, where $m=-3$ and $b=-3$. Notice that

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

and

$$\lim_{x\to-\infty}\frac{-3x^2+4}{x-1}-(-3x-3)=\lim_{x\to-\infty}\frac{1}{x-1}=0.$$

Thus, $y=-3x-3$ is a slant asymptote of $f(x)$. ♣

Although rational functions are the most common type of function we encounter with slant asymptotes, there are other types of functions we can consider that present an interesting challenge.

Example 3.26: Slant Asymptote

Show that $y=2x+4$ *is a slant asymptote of* $f(x)=2x-3^x+4$.

Solution. This is because

$$\lim_{x\to-\infty}[f(x)-(2x+4)]=\lim_{x\to-\infty}(-3^x)=0.$$

♣

We note that $\lim_{x\to\infty}[f(x)-(2x+4)]=\lim_{x\to\infty}(-3^x)=-\infty$. So, the vertical distance between $y=f(x)$ and the line $y=2x+4$ decreases toward 0 only when $x\to-\infty$ and not when $x\to\infty$. The graph of f

approaches the slant asymptote $y = 2x + 4$ only at the far left and not at the far right. One might ask if $y = f(x)$ approaches a slant asymptote when $x \to \infty$. The answer turns out to be no, but we will need to know something about the relative growth rates of the exponential functions and linear functions in order to prove this. Specifically, one can prove that when the base is greater than 1 the exponential functions grows faster than any power function as $x \to \infty$. This can be phrased like this: For any $a > 1$ and any $n > 0$,

$$\lim_{x\to\infty} \frac{a^x}{x^n} = \infty \text{ and } \lim_{x\to\infty} \frac{x^n}{a^x} = 0.$$

These facts are most easily proved with the aim of something called the L'Hôpital's Rule.

3.5.4. End Behaviour and Comparative Growth Rates

Let us now look at the last two subsections and go deeper. In the last two subsections we looked at horizontal and slant asymptotes. Both are special cases of the end behaviour of functions, and both concern situations where the graph of a function approaches a straight line as $x \to \infty$ or $-\infty$. But not all functions have this kind of end behviour. For example, $f(x) = x^2$ and $f(x) = x^3$ do not approach a straight line as $x \to \infty$ or $-\infty$. The best we can say with the notion of limit developed at this stage are that

$$\begin{aligned} \lim_{x\to\infty} x^2 &= \infty, \lim_{x\to-\infty} x^2 = \infty, \\ \lim_{x\to\infty} x^3 &= \infty, \lim_{x\to-\infty} x^3 = -\infty. \end{aligned}$$

Similarly, we can describe the end behaviour of transcendental functions such as $f(x) = e^x$ using limits, and in this case, the graph approaches a line as $x \to -\infty$ but not as $x \to \infty$.

$$\lim_{x\to-\infty} e^x = 0, \lim_{x\to\infty} e^x = \infty.$$

People have found it useful to make a finer distinction between these end behaviours all thus far captured by the symbols ∞ and $-\infty$. Specifically, we will see that the above functions have different growth rates at infinity. Some increases to infinty faster than others. Specifically,

Definition 3.27: Comparative Growth Rates

Suppose that f and g are two functions such that $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to\infty} g(x) = \infty$. We say that $f(x)$ grows faster than $g(x)$ as $x \to \infty$ if the following holds:

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

or equivalently,

$$\lim_{x\to\infty} \frac{g(x)}{f(x)} = 0.$$

Here are a few obvious examples:

Example 3.28:

Show that if $m > n$ are two positive integers, then $f(x) = x^m$ grows faster than $g(x) = x^n$ as $x \to \infty$.

Solution. Since $m > n$, $m - n$ is a positive integer. Therefore,

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{x^m}{x^n} = \lim_{x\to\infty} x^{m-n} = \infty.$$

♣

Example 3.29:

Show that if $m > n$ are two positive integers, then any monic polynomial $P_m(x)$ of degree m grows faster than any monic polynomial $P_n(x)$ of degree n as $x \to \infty$. [Recall that a polynomial is monic if its leading coefficient is 1.]

Solution. By assumption, $P_m(x) = x^m +$ terms of degrees less than $m = x^m + a_{m-1}x^{m-1} + \ldots$, and $P_n(x) = x^n +$ terms of degrees less than $n = x^n + b_{n-1}x^{n-1} + \ldots$. Dividing the numerator and denominator by x^n, we get

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{x^{m-n} + a_{m-1}x^{m-n-1} + \ldots}{1 + \frac{b_{n-1}}{x} + \ldots} = \lim_{x\to\infty} x^{m-n}\left(\frac{1 + \frac{a_{m-1}}{x} + \ldots}{1 + \frac{b_{n-1}}{x} + \ldots}\right) = \infty,$$

since the limit of the bracketed fraction is 1 and the limit of x^{m-n} is ∞, as we showed in Example 3.28.

♣

Example 3.30:

Show that a polynomial grows exactly as fast as its highest degree term as $x \to \infty$ or $-\infty$. That is, if $P(x)$ is any polynomial and $Q(x)$ is its highest degree term, then both limits

$$\lim_{x\to\infty} \frac{P(x)}{Q(x)} \text{ and } \lim_{x\to-\infty} \frac{P(x)}{Q(x)}$$

are finite and nonzero.

Solution. Suppose that $P(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$, where $a_n \neq 0$. Then the highest degree term is $Q(x) = a_n x^n$. So,

$$\lim_{x\to\infty} \frac{P(x)}{Q(x)} = \lim_{x\to\infty} \left(a_n + \frac{a_{n-1}}{x} + \ldots + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}\right) = a_n \neq 0.$$

♣

Let's state a theorem we mentioned when we discussed the last example in the last subsection:

Theorem 3.31:

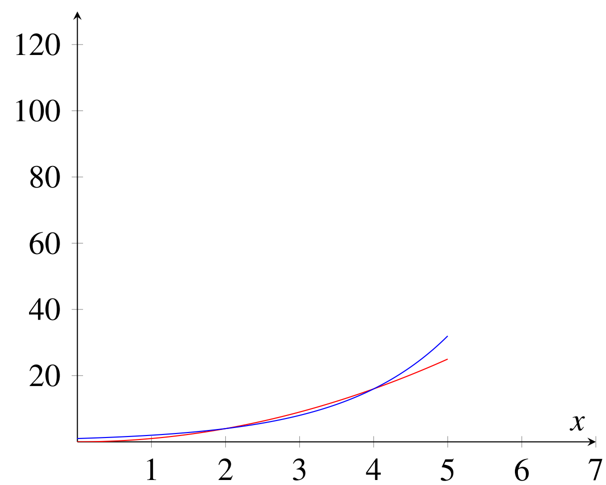
Let n be any positive integer and let $a > 1$. Then $f(x) = a^x$ grows faster than $g(x) = x^n$ as $x \to \infty$:

$$\lim_{x\to\infty} \frac{a^x}{x^n} = \infty, \ \lim_{x\to\infty} \frac{x^n}{a^x} = 0.$$

In particular,

$$\lim_{x\to\infty} \frac{e^x}{x^n} = \infty, \ \lim_{x\to\infty} \frac{x^n}{e^x} = 0.$$

The easiest way to prove this is to use the L'Hôpital's Rule, which we will introduce in a later chapter. For now, one can plot and compare the graphs of an exponential function and a power function. Here is a comparison between $f(x) = x^2$ and $g(x) = 2^x$:



Notice also that as $x \to -\infty$, x^n grows in size but e^x does not. More specifically, $x^n \to \infty$ or $-\infty$ according as n is even or odd, while $e^x \to 0$. So, it is meaningless to compare their "growth" rates, although we can still calculate the limit

$$\lim_{x\to-\infty} \frac{e^x}{x^n} = 0.$$

Let's see an application of our theorem.

Example 3.32:

Find the horizontal asymptote(s) of $f(x) = \dfrac{x^3 + 2e^x}{e^x - 4x^2}$.

Solution. To find horizontal asymptotes, we calculate the limits of $f(x)$ as $x \to \infty$ and $x \to -\infty$. For $x \to \infty$, we divide the numerator and the denominator by e^x, and then we take limit to get

$$\lim_{x\to\infty} \frac{x^3 + 2e^x}{e^x - 4x^2} = \lim_{x\to\infty} \frac{\frac{x^3}{e^x} + 2}{1 - 4\frac{x^2}{e^x}} = \frac{0+2}{1-4(0)} = 2.$$

For $x \to -\infty$, we divide the numerator and the denominator by x^2 to get

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

The denominator now approaches $0-4=-4$. The numerator has limit $-\infty$. So, the quotient has limit ∞:

$$\lim_{x\to-\infty} \frac{x+2\frac{e^x}{x^2}}{\frac{e^x}{x^2}-4} = \infty.$$

So, $y=2$ is a horizontal asymptote. The function $y=f(x)$ approaches the line $y=2$ as $x\to\infty$. And this is the only horizontal asymptote, since the function $y=f(x)$ does not approach any horizontal line as $x\to-\infty$. ♣

Since the growth rate of a polynomial is the same as that of its leading term, the following is obvious:

Example 3.33:

If $P(x)$ is any polynomial, then

$$\lim_{x\to\infty} \frac{P(x)}{e^x} = 0.$$

Also, if r is any real number, then we can place it between two consecutive integers n and $n+1$. For example, $\sqrt{3}$ is between 1 and 2, e is between 2 and 3, and π is between 3 and 4. Then the following is totally within our expectation:

Example 3.34:

Prove that if $a>1$ is any basis and $r>0$ is any exponent, then $f(x)=a^x$ grows faster than $g(x)=x^r$ as $x\to\infty$.

Solution. Let r be between consecutive integers n and $n+1$. Then for all $x>1$, $x^n \le x^r \le x^{n+1}$. Dividing by a^x, we get

$$\frac{x^n}{a^x} \le \frac{x^r}{a^x} \le \frac{x^{n+1}}{a^x}.$$

Since

$$\lim_{x\to\infty} \frac{x^n}{a^x} = 0.$$

♣

What about exponential functions with different bases? We recall from the graphs of the exponential functions that for any base $a>1$,

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

So, the exponential functions with bases greater than 1 all grow to infinity as $x\to\infty$. How do their growth rates compare?

Theorem 3.35:

If $1 < a < b$, then $f(x) = b^x$ grows faster than $g(x) = a^x$ as $x \to \infty$.

Proof. Proof. Since $a < b$, we have $\frac{b}{a} > 1$. So,

$$\lim_{x\to\infty} \frac{b^x}{a^x} = \lim_{x\to\infty} \left(\frac{b}{a}\right)^x = \infty.$$

♣

Another function that grows to infinity as $x \to \infty$ is $g(x) = \ln x$. Recall that the natural logarithmic function is the inverse of the exponential function $y = e^x$. Since e^x grows very fast as x increases, we should expect $\ln x$ to grow very slowly as x increases. The same applies to logarithmic functions with any basis $a > 1$. This is the content of the next theorem.

Theorem 3.36:

Let r be any positive real number and $a > 1$. Then

(a) $f(x) = x^r$ grows faster than $g(x) = \ln x$ as $x \to \infty$.

(b) $f(x) = x^r$ grows faster than $g(x) = \log_a x$ as $x \to \infty$.

Proof.

1. We use a change of variable. Letting $t = \ln x$, then $x = e^t$. So, $x \to \infty$ if and only if $t \to \infty$, and

$$\lim_{x\to\infty} \frac{\ln x}{x^r} = \lim_{t\to\infty} \frac{t}{(e^t)^r} = \lim_{t\to\infty} \frac{t}{(e^r)^t}.$$

Now, since $r > 0$, $a = e^r > 1$. So, a^t grows as t increases, and it grows faster than t as $t \to \infty$. Therefore,

$$\lim_{x\to\infty} \frac{\ln x}{x^r} = \lim_{t\to\infty} \frac{t}{(e^r)^t} = \lim_{t\to\infty} \frac{t}{a^t} = 0.$$

2. The change of base identity $\log_a x = \dfrac{\ln x}{\ln a}$ implies that $\log_a x$ is simply a constant multple of $\ln x$. The result now follows from (a).

♣

Exercises for 3.5

Exercise 3.5.1 *Compute the following limits.*

(a) $\lim_{x\to\infty} \sqrt{x^2+x}-\sqrt{x^2-x}$

(b) $\lim_{x\to\infty} \frac{e^x+e^{-x}}{e^x-e^{-x}}$

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

(d) $\lim_{t\to\infty} \frac{t+5-2/t-1/t^3}{3t+12-1/t^2}$

(e) $\lim_{y\to\infty} \frac{\sqrt{y+1}+\sqrt{y-1}}{y}$

(f) $\lim_{x\to 0^+} \frac{3+x^{-1/2}+x^{-1}}{2+4x^{-1/2}}$

(g) $\lim_{x\to\infty} \frac{x+x^{1/2}+x^{1/3}}{x^{2/3}+x^{1/4}}$

(h) $\lim_{t\to\infty} \frac{1-\sqrt{\frac{t}{t+1}}}{2-\sqrt{\frac{4t+1}{t+2}}}$

(i) $\lim_{t\to\infty} \frac{1-\frac{t}{t-1}}{1-\sqrt{\frac{t}{t-1}}}$

(j) $\lim_{x\to -\infty} \frac{x+x^{-1}}{1+\sqrt{1-x}}$

(k) $\lim_{x\to 1^+} \frac{\sqrt{x}}{x-1}$

(l) $\lim_{x\to\infty} \frac{x^{-1}+x^{-1/2}}{x+x^{-1/2}}$

(m) $\lim_{x\to\infty} \frac{x+x^{-2}}{2x+x^{-2}}$

(n) $\lim_{x\to\infty} \frac{5+x^{-1}}{1+2x^{-1}}$

(o) $\lim_{x\to\infty} \frac{4x}{\sqrt{2x^2+1}}$

(p) $\lim_{x\to\infty} (x+5)\left(\frac{1}{2x}+\frac{1}{x+2}\right)$

(q) $\lim_{x\to 0^+} (x+5)\left(\frac{1}{2x}+\frac{1}{x+2}\right)$

(r) $\lim_{x\to 2} \frac{x^3-6x-2}{x^3-4x}$

Exercise 3.5.2 *The function* $f(x)=\frac{x}{\sqrt{x^2+1}}$ *has two horizontal asymptotes. Find them and give a rough sketch of* f *with its horizontal asymptotes.*

Exercise 3.5.3 *Find the vertical asymptotes of* $f(x)=\frac{\ln x}{x-2}$.

Exercise 3.5.4 *Suppose that a falling object reaches velocity* $v(t)=50(1-e^{-t/5})$ *at time* t*, where distance is measured in m and time s. What is the object's terminal velocity, i.e. the value of* $v(t)$ *as* t *goes to infinity?*

Exercise 3.5.5 *Find the slant asymptote of* $f(x)=\frac{x^2+x+6}{x-3}$.

Exercise 3.5.6 *Compute the following limits.*

(a) $\lim_{x\to -\infty} (2x^3-x)$

(b) $\lim_{x\to\infty} \tan^{-1}(e^x)$

(c) $\lim_{x\to -\infty} \tan^{-1}(e^x)$

(d) $\lim_{x\to\infty} \frac{e^x+x^4}{x^3+5\ln x}$

(e) $\lim_{x\to\infty} \frac{2^x+5(3^x)}{3(2^x)-3^x}$

(f) $\lim_{x\to -\infty} \frac{2^x+5(3^x)}{3(2^x)-3^x}$

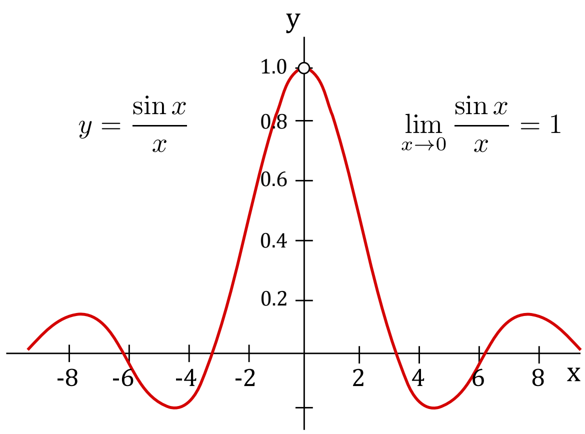
(g) $\lim_{x \to 0^+} \sqrt{x} \ln x$ [Hint: Let $t = 1/x$]

3.6 A Trigonometric Limit

In this section we aim to compute the limit:

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

We start by analyzing the graph of $y = \dfrac{\sin x}{x}$:



Notice that $x = 0$ is not in the domain of this function. Nevertheless, we can look at the limit as x approaches 0. From the graph we find that the limit is 1 (there is an open circle at $x = 0$ indicating 0 is not in the domain). We just convinced you this limit formula holds true based on the graph, but how does one attempt to prove this limit more formally? To do this we need to be quite clever, and to employ some indirect reasoning. The indirect reasoning is embodied in a theorem, frequently called the **squeeze theorem**.

Theorem 3.37: Squeeze Theorem

Suppose that $g(x) \leq f(x) \leq h(x)$ for all x close to a but not equal to a. If $\lim_{x \to a} g(x) = L = \lim_{x \to a} h(x)$, then $\lim_{x \to a} f(x) = L$.

This theorem can be proved using the official definition of limit. We won't prove it here, but point out that it is easy to understand and believe graphically. The condition says that $f(x)$ is trapped between $g(x)$ below and $h(x)$ above, and that at $x = a$, both g and h approach the same value. This means the situation looks something like Figure 3.1.

For example, imagine the blue curve is $f(x) = x^2 \sin(\pi/x)$, the upper (red) and lower (green) curves are $h(x) = x^2$ and $g(x) = -x^2$. Since the sine function is always between -1 and 1, $-x^2 \leq x^2 \sin(\pi/x) \leq x^2$, and it is easy to see that $\lim_{x\to 0} -x^2 = 0 = \lim_{x\to 0} x^2$. It is not so easy to see directly (i.e. algebraically) that $\lim_{x\to 0} x^2 \sin(\pi/x) = 0$, because the π/x prevents us from simply plugging in $x = 0$. The squeeze theorem makes this "hard limit" as easy as the trivial limits involving x^2.

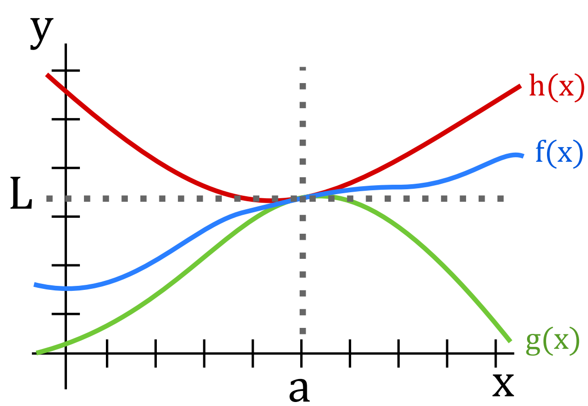


Figure 3.1: The squeeze theorem.

To compute $\lim\limits_{x\to 0}(\sin x)/x$, we will find two simpler functions g and h so that $g(x) \leq (\sin x)/x \leq h(x)$, and so that $\lim_{x\to 0} g(x) = \lim_{x\to 0} h(x)$. Not too surprisingly, this will require some trigonometry and geometry. Referring to figure 3.2, x is the measure of the angle in radians. Since the circle has radius 1, the coordinates of point A are $(\cos x, \sin x)$, and the area of the small triangle is $(\cos x \sin x)/2$. This triangle is completely contained within the circular wedge-shaped region bordered by two lines and the circle from $(1,0)$ to point A. Comparing the areas of the triangle and the wedge we see $(\cos x \sin x)/2 \leq x/2$, since the area of a circular region with angle θ and radius r is $\theta r^2/2$. With a little algebra this turns into $(\sin x)/x \leq 1/\cos x$, giving us the h we seek.

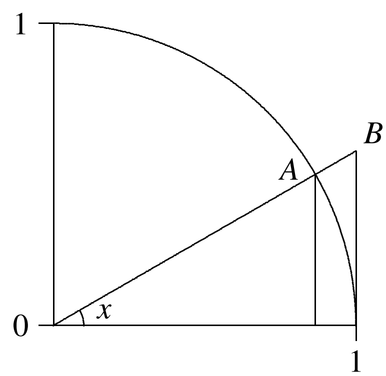


Figure 3.2: Visualizing $\sin x/x$**.**

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

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

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

Using the above, we can compute a similar limit:

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

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

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

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

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

Thus,

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

Example 3.38: Limit of Other Trig Functions

Compute the following limit $\lim_{x\to 0}\frac{\sin 5x\cos x}{x}$.

Solution. We have

$$\begin{aligned}\lim_{x\to 0}\frac{\sin 5x\cos x}{x} &= \lim_{x\to 0}\frac{5\sin 5x\cos x}{5x}\\ &= \lim_{x\to 0}5\cos x\left(\frac{\sin 5x}{5x}\right)\\ &= 5\cdot(1)\cdot(1) \;=\; 5\end{aligned}$$

since $\cos(0)=1$ and $\lim_{x\to 0}\frac{\sin 5x}{5x}=1$. ♣

Let's do a harder one now.

Example 3.39: Limit of Other Trig Functions

Compute the following limit: $\lim_{x\to 0}\frac{\tan^3 2x}{x^2\sin 7x}$.

Solution. Recall that the $\tan^3(2x)$ means that $\tan(2x)$ is being raised to the third power.

$$\begin{aligned}
\lim_{x\to 0} \frac{\tan^3(2x)}{x^2\sin(7x)} &= \lim_{x\to 0} \frac{(\sin(2x))^3}{x^2\sin(7x)\cos^3(2x)} && \text{Rewrite in terms of sin and cos} \\
&= \lim_{x\to 0} \frac{(2x)^3\left(\frac{\sin(2x)}{2x}\right)^3}{x^2(7x)\left(\frac{\sin(7x)}{7x}\right)\cos^3(2x)} && \text{Make sine terms look like: } \frac{\sin\theta}{\theta} \\
&= \lim_{x\to 0} \frac{8x^3(1)^3}{7x^3(1)(1^3)} && \text{Replace } \lim_{x\to 0}\frac{\sin nx}{nx} \text{ with 1. Also, } \cos(0)=1. \\
&= \lim_{x\to 0} \frac{8}{7} && \text{Cancel } x^3\text{'s.} \\
&= \frac{8}{7}.
\end{aligned}$$

♣

Example 3.40: Applying the Squeeze Theorem

Compute the following limit: $\displaystyle\lim_{x\to 0^+} x^3\cos\left(\frac{1}{\sqrt{x}}\right)$.

Solution. We use the *Squeeze Theorem* to evaluate this limit. We know that $\cos\alpha$ satisfies $-1 \le \cos\alpha \le 1$ for any choice of α. Therefore we can write:

$$-1 \le \cos\left(\frac{1}{\sqrt{x}}\right) \le 1$$

Since $x \to 0^+$ implies $x > 0$, multiplying by x^3 gives:

$$-x^3 \le x^3\cos\left(\frac{1}{\sqrt{x}}\right) \le x^3.$$

$$\lim_{x\to 0^+}(-x^3) \le \lim_{x\to 0^+}\left(x^3\cos\left(\frac{1}{\sqrt{x}}\right)\right) \le \lim_{x\to 0^+} x^3.$$

But using our rules we know that

$$\lim_{x\to 0^+}(-x^3) = 0, \qquad \lim_{x\to 0^+} x^3 = 0$$

and the Squeeze Theorem says that the only way this can happen is if

$$\lim_{x\to 0^+} x^3\cos\left(\frac{1}{\sqrt{x}}\right) = 0.$$

♣

When solving problems using the Squeeze Theorem it is also helpful to have the following theorem.

Theorem 3.41: Monotone Limits

If $f(x) \leq g(x)$ when x is near a (except possibly at a) and the limits of f and g both exist as x approaches a, then $\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x)$.

Exercises for 3.6

Exercise 3.6.1 *Compute the following limits.*

(a) $\displaystyle\lim_{x\to 0} \frac{\sin(5x)}{x}$

(b) $\displaystyle\lim_{x\to 0} \frac{\sin(7x)}{\sin(2x)}$

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

(d) $\displaystyle\lim_{x\to 0} \frac{\tan x}{x}$

(e) $\displaystyle\lim_{x\to \pi/4} \frac{\sin x - \cos x}{\cos(2x)}$

Exercise 3.6.2 *For all $x \geq 0$, $4x-9 \leq f(x) \leq x^2-4x+7$. Find $\lim_{x\to 4} f(x)$.*

Exercise 3.6.3 *For all x, $2x \leq g(x) \leq x^4-x^2+2$. Find $\lim_{x\to 1} g(x)$.*

Exercise 3.6.4 *Use the Squeeze Theorem to show that $\lim_{x\to 0} x^4\cos(2/x) = 0$.*

Exercise 3.6.5 *Find the value of $\lim_{x\to\infty} \dfrac{3x+\sin x}{x+\cos x}$. Justify your steps carefully.*

3.7 Continuity

The graph shown in Figure 3.3(a) represents a **continuous** function. Geometrically, this is because there are no jumps in the graphs. That is, if you pick a point on the graph and approach it from the left and right, the values of the function approach the value of the function at that point. For example, we can see that this is not true for function values near $x = 1$ on the graph in Figure 3.3(b) which is not continuous at that location.

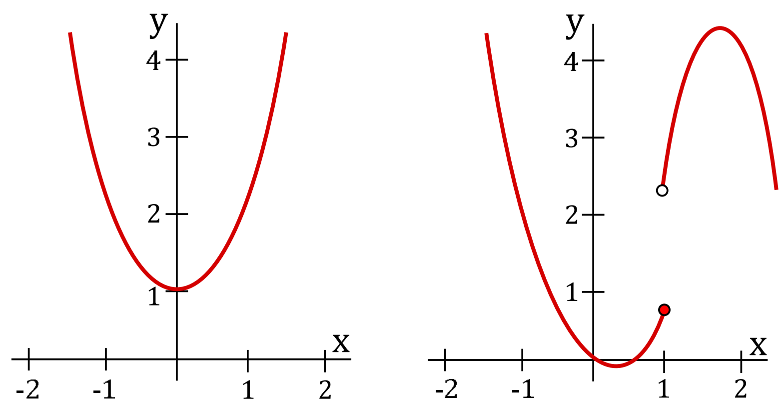


Figure 3.3: (a) A continuous function. (b) A function with a discontinuity at $x = 1$.

Definition 3.42: Continuous at a Point

A function f is **continuous at a point** a *if*

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

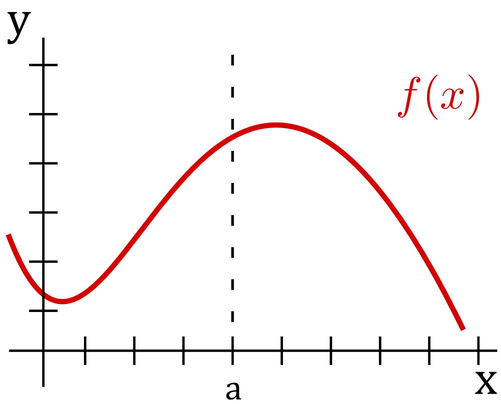
Some readers may prefer to think of continuity at a point as a three part definition. That is, a function $f(x)$ is continuous at $x = a$ if the following three conditions hold:

(i) $f(a)$ is defined (that is, a belongs to the domain of f),

(ii) $\lim\limits_{x \to a} f(x)$ exists (that is, left-hand limit $=$ right-hand limit),

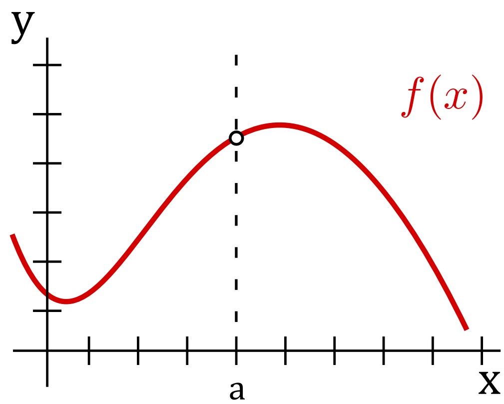
(iii) $\lim\limits_{x \to a} f(x) = f(a)$ (that is, the numbers from (i) and (ii) are equal).

The figures below show graphical examples of functions where either (i), (ii) or (iii) can fail to hold.



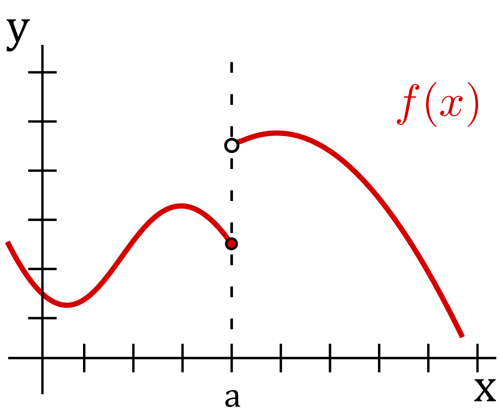
continuous at $x = a$

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



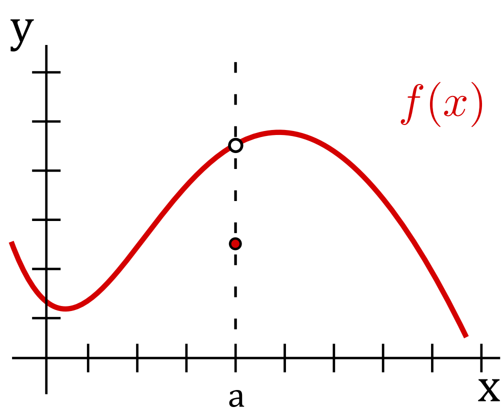
f(a) not defined

(i) fails to hold



$\lim_{x \to a} f(x)$ does not exist

(ii) fails to hold



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

(iii) fails to hold

On the other hand, if f is defined on an open interval containing a, except perhaps at a, we can say that f is **discontinuous** at a if f is not continuous at a.

Graphically, you can think of continuity as being able to draw your function without having to lift your pencil off the paper. If your pencil has to jump off the page to continue drawing the function, then the function is not continuous at that point. This is illustrated in Figure 3.3(b) where if we tried to draw the function (from left to right) we need to lift our pencil off the page once we reach the point $x = 1$ in order to be able to continue drawing the function.

Definition 3.43: Continuity on an Open Interval

A function f is **continuous on an open interval** (a,b) *if it is continuous at every point in the interval.*

Furthermore, a function is **everywhere continuous** *if it is continuous on the entire real number line* $(-\infty,\infty)$..

Recall the function graphed in a previous section as shown in Figure 3.4.

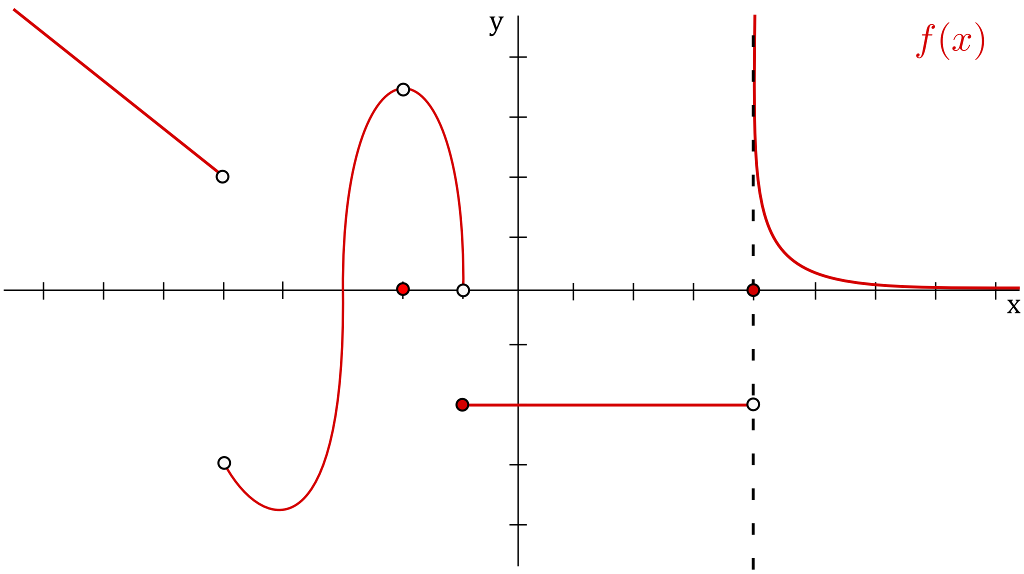
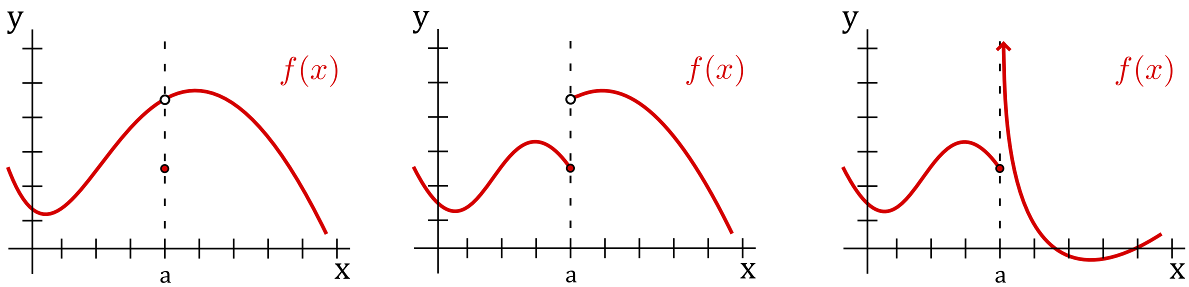


Figure 3.4: A function with discontinuities at $x=-5$, $x=-2$, $x=-1$ and $x=4$.

We can draw this function without lifting our pencil *except* at the points $x=-5$, $x=-2$, $x=-1$, and $x=4$. Thus, $f(x)$ is *continuous* at every real number *except* at these four numbers. At $x=-5$, $x=-2$, $x=-1$, and $x=4$, the function $f(x)$ is *discontinuous*.

At $x=-2$ we have a **removable discontinuity** because we could remove this discontinuity simply by redefining $f(-2)$ to be 3.5. At $x=-5$ and $x=-1$ we have **jump discontinuities** because the function jumps from one value to another. From the right of $x=4$, we have an **infinite discontinuity** because the function goes off to infinity.

Formally, we say $f(x)$ has a **removable discontinuity** at $x=a$ if $\lim_{x\to a} f(x)$ exists but is not equal to $f(a)$. Note that we do not require $f(a)$ to be defined in this case, that is, a need not belong to the domain of $f(x)$.



$\lim\limits_{x \to a} f(x) \neq f(a)$ — REMOVABLE DISCONTINUITY

$\lim\limits_{x \to a^+} f(x) \neq \lim\limits_{x \to a^-} f(x)$ — JUMP DISCONTINUITY

Either $\lim\limits_{x \to a^+} f(x) = \pm\infty$ or $\lim\limits_{x \to a^-} f(x) = \pm\infty$ — INFINITE DISCONTINUITY

Example 3.44: Continuous at a Point

What value of c will make the following function $f(x)$ continuous at 2?

$$f(x) = \begin{cases} \dfrac{x^2 - x - 2}{x - 2} & \text{if } x \neq 2 \\ c & \text{if } x = 2 \end{cases}$$

Solution. In order to be continuous at 2 we require

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

to hold. We use the three part definition listed previously to check this.

1. First, $f(2) = c$, and c is some real number. Thus, $f(2)$ is defined.

2. Now, we must evaluate the limit. Rather than computing both one-sided limits, we just compute the limit directly. For x close to 2 (but not equal to 2) we can replace $f(x)$ with $\frac{x^2-x-2}{x-2}$ to get:

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

Therefore the limit exists and equals 3.

3. Finally, for f to be continuous at 2, we need that the numbers in the first two items to be equal. Therefore, we require $c = 3$. Thus, when $c = 3$, $f(x)$ is continuous at 2, for any other value of c, $f(x)$ is discontinuous at 2. ♣

For continuity on a closed interval, we consider the **one-sided limits** of a function. Recall that $x \to a^-$ means x approaches a from values less than a.

Definition 3.45: Continuous from the Right and from the Left

A function f is **left continuous at a point a** *if*

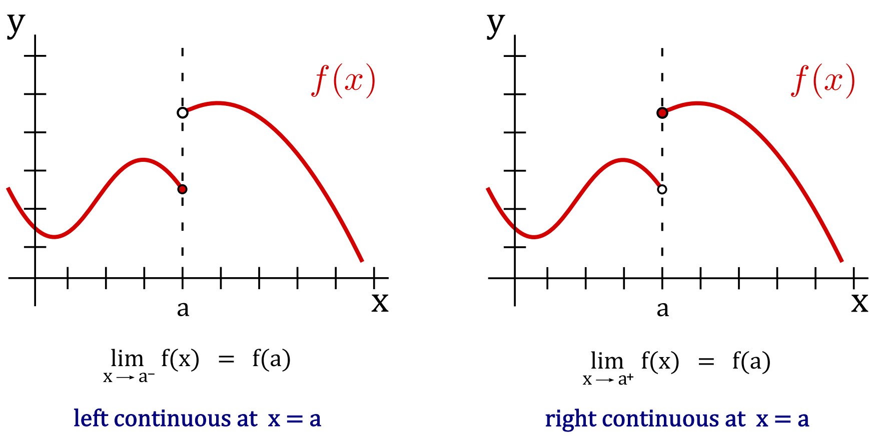
$$\lim_{x\to a^-} f(x) = f(a)$$

and **right continuous at a point a** *if*

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

If a function f is continuous at a, then it is both left and right continuous at a.

The above definition regarding left (or right) continuous functions is illustrated with the following figure:



One-sided limits allows us to extend the definition of continuity to closed intervals. The following definition means a function is continuous on a closed interval if it is continuous in the interior of the interval and possesses the appropriate one-sided continuity at the endpoints of the interval.

Definition 3.46: Continuity on a Closed Interval

A function f is **continuous on the closed interval** $[a,b]$ *if:*

(i) it is continuous on the open interval (a,b);

(ii) it is left continuous at point a:

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

and

(iii) it is right continuous at point b:

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

This definition can be extended to continuity on half-open intervals such as $(a,b]$ and $[a,b)$, and unbounded intervals.

Example 3.47: Continuity on Other Intervals

The function $f(x) = \sqrt{x}$ *is continuous on the (closed) interval* $[0,\infty)$.

The function $f(x) = \sqrt{4-x}$ *is continuous on the (closed) interval* $(-\infty,4]$.

The continuity of functions is preserved under the operations of addition, subtraction, multiplication and division (in the case that the function in the denominator is nonzero).

Theorem 3.48: Operations of Continuous Functions

If f *and* g *are continous at* a*, and* c *is a constant, then the following functions are also continuous at* a*:*

(i) $f \pm g$;

(ii) cf;

(iii) fg;

(iv) f/g *(provided* $g(a) \neq 0$*).*

Below we list some common functions that are known to be continuous on every interval inside their domains.

Example 3.49: Common Types of Continuous Functions

- *Polynomials (for all* x*), e.g.,* $y = mx+b$, $y = ax^2+bx+c$.
- *Rational functions (except at points* x *which gives division by zero).*
- *Root functions* $\sqrt[n]{x}$ *(for all* x *if* n *is odd, and for* $x \geq 0$ *if* n *is even).*
- *Trigonometric functions*
- *Inverse trigonometric functions*
- *Exponential funtions*
- *Logarithmic functions*

For rational functions with removable discontinuities as a result of a zero, we can define a new function filling in these gaps to create a piecewise function that is continuous everywhere.

Continuous functions are where the *direct substitution property* hold. This fact can often be used to compute the limit of a continuous function.

Example 3.50: Evaluate a Limit

Evaluate the following limit: $\lim\limits_{x\to\pi} \dfrac{\sqrt{x}+\sin x}{1+x+\cos x}$.

Solution. We will use a continuity argument to justify that direct substitution can be applied. By the list above, $\sqrt{x}$, $\sin x$, 1, x and $\cos x$ are all continuous functions at π. Then $\sqrt{x}+\sin x$ and $1+x+\cos x$ are both continuous at π. Finally,

$$\frac{\sqrt{x}+\sin x}{1+x+\cos x}$$

is a continuous function at π since $1+\pi+\cos\pi \neq 0$. Hence, we can directly substitute to get the limit:

$$\lim_{x\to\pi}\frac{\sqrt{x}+\sin x}{1+x+\cos x}=\frac{\sqrt{\pi}+\sin\pi}{1+\pi+\cos\pi}=\frac{\sqrt{\pi}}{\pi}=\frac{1}{\sqrt{\pi}}.$$

♣

Continuity is also preserved under the composition of functions.

Theorem 3.51: Continuity of Function Composition

If g is continuous at a and f is continuous at $g(a)$, then the composition function $f\circ g$ is continuous at a.

Example 3.52: Continuity with Composition of Functions

Determine where the following functions is continuous:

(a) $h(x)=\cos(x^2)$

(b) $H(x)=\ln(1+\sin(x))$

Solution.

(a) The functions that make up the composition $h(x)=f(g(x))$ are $g(x)=x^2$ and $f(x)=\cos(x)$. The function g is continuous on $\mathbb{R}$ since it is a polynomial, and f is also continuous everywhere. Therefore, $h(x)=(f\circ g)(x)$ is continuous on $\mathbb{R}$ by Theorem 3.51.

(b) We know from Example 3.49 that $f(x)=\ln x$ is continuous and $g(x)=1+\sin x$ are continuous. Thus by Theorem 3.51, $H(x)=f(g(x))$ is continuous wherever it is defined. Now $\ln(1+\sin x)$ is defined when $1+\sin x>0$. Recall that $-1\leq\sin x\leq 1$, so $1+\sin x>0$ except when $\sin x=-1$, which happens when $x=\pm3\pi/2,\pm7\pi/2,\ldots$. Therefore, H has discontinuities when $x=3\pi n/2$, $n=1,2,3,\ldots$ and is continuous on the intervals between these values.

♣

Intermediate Value Theorem

Whether or not an equation *has* a solution is an important question in mathematics. Consider the following two questions:

Example 3.53: Motivation for the Intermediate Value Theorem

1. *Does* $e^x + x^2 = 0$ *have a solution?*

2. *Does* $e^x + x = 0$ *have a solution?*

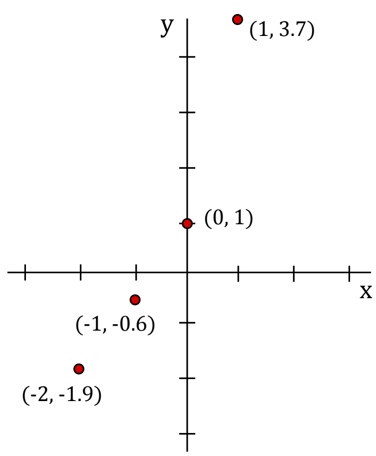
Solution.

1. The first question is easy to answer since for any exponential function we know that $a^x > 0$, and we also know that whenever you square a number you get a nonnegative answer: $x^2 \geq 0$. Hence, $e^x + x^2 > 0$, and thus, is never equal to zero. Therefore, the first equation has no solution.

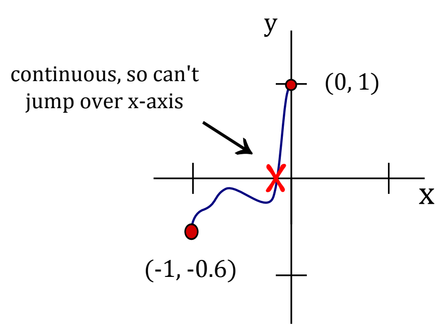
2. For the second question, it is difficult to see if $e^x + x = 0$ has a solution. If we tried to solve for x, we would run into problems. Let's make a table of values to see what kind of values we get (recall that $e \approx 2.7183$):

x	$e^x + x$
-2	$e^{-2} - 2 \approx -1.9$
-1	$e^{-1} - 1 \approx -0.6$
0	$e^0 + 0 = 1$
1	$e + 1 \approx 3.7$

Sketching this gives:



Let $f(x) = e^x + x$. Notice that if we choose $a = -1$ and $b = 0$ then we have $f(a) < 0$ and $f(b) > 0$. A point where the function $f(x)$ crosses the x-axis gives a solution to $e^x + x = 0$. Since $f(x) = e^x + x$ is continuous (both e^x and x are continuous), then the function *must* cross the x-axis somewhere between -1 and 0:



Therefore, our equation has a solution.

Note that by looking at smaller and smaller intervals (a,b) with $f(a) < 0$ and $f(b) > 0$, we can get a better and better approximation for a solution to $e^x + x = 0$. For example, taking the interval $(-0.4, -0.6)$ gives $f(-0.4) < 0$ and $f(-0.6) > 0$, thus, there is a solution to $f(x) = 0$ between -0.4 and -0.6. It turns out that the solution to $e^x + x = 0$ is $x \approx -0.56714$.

♣

We now generalize the argument used in the previous example. In that example we had a continuous function that went from negative to positive and hence, had to cross the x-axis at some point. In fact, we don't need to use the x-axis, any line $y = N$ will work so long as the function is continuous and below the line $y = N$ at some point and above the line $y = N$ at another point. This is known as the Intermediate Values Theorem and it is formally stated as follows:

Theorem 3.54: Intermediate Value Theorem

If f is continuous on the interval $[a,b]$ and N is between $f(a)$ and $f(b)$, where $f(a) \neq f(b)$, then there is a number c in (a,b) such that $f(c) = N$.

The Intermediate Value Theorem guarantees that if $f(x)$ is continuous and $f(a) < N < f(b)$, the line $y = N$ intersects the function at some point $x = c$. Such a number c is between a and b and has the property that $f(c) = N$ (see Figure 3.5(a)). We can also think of the theorem as saying if we draw the line $y = N$ between the lines $y = f(a)$ and $y = f(b)$, then the function cannot jump over the line $y = N$. On the other hand, if $f(x)$ is *not* continuous, then the theorem may *not* hold. See Figure 3.5(b) where there is no number c in (a,b) such that $f(c) = N$. Finally, we remark that there may be multiple choices for c (i.e., lots of numbers between a and b with y-coordinate N). See Figure 3.5(c) for such an example.

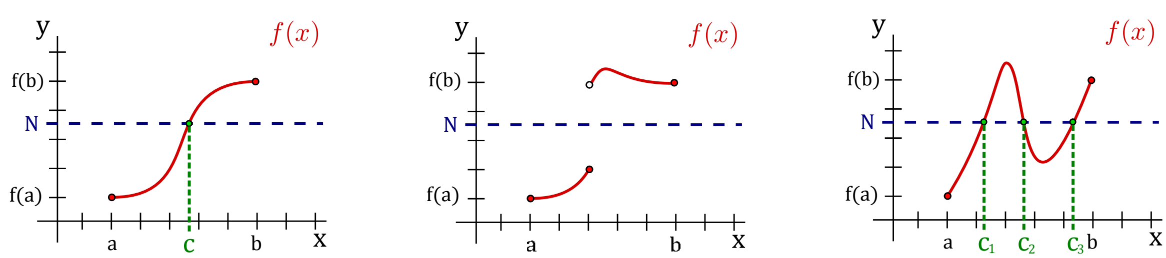


Figure 3.5: (a) A continuous function where IVT holds for a single value c. (b) A discontinuous function where IVT fails to hold. (c) A continuous function where IVT holds for multiple values in (a,b).

The Intermediate Value Theorem is most frequently used for $N = 0$.

Example 3.55: Intermediate Value Theorem

Show that there is a solution of $\sqrt[3]{x} + x = 1$ in the interval $(0,8)$.

Solution. Let $f(x) = \sqrt[3]{x} + x - 1$, $N = 0$, $a = 0$, and $b = 8$. Since $\sqrt[3]{x}$, x and -1 are continuous on $\mathbb{R}$, and the sum of continuous functions is again continuous, we have that $f(x)$ is continuous on $\mathbb{R}$, thus in particular, $f(x)$ is continuous on $[0,8]$. We have $f(a) = f(0) = \sqrt[3]{0} + 0 - 1 = -1$ and $f(b) = f(8) = \sqrt[3]{8} + 8 - 1 = 9$.

Thus $N=0$ lies between $f(a)=-1$ and $f(b)=9$, so the conditions of the Intermediate Value Theorem are satisfied. So, there exists a number c in $(0,8)$ such that $f(c)=0$. This means that c satisfies $\sqrt[3]{c}+c-1=0$, in otherwords, is a solution for the equation given.

Alternatively we can let $f(x)=\sqrt[3]{x}+x$, $N=1$, $a=0$ and $b=8$. Then as before $f(x)$ is the sum of two continuous functions, so is also continuous everywhere, in particular, continuous on the interval $[0,8]$. We have $f(a)=f(0)=\sqrt[3]{0}+0=0$ and $f(b)=f(8)=\sqrt[3]{8}+8=10$. Thus $N=1$ lies between $f(a)=0$ and $f(b)=10$, so the conditions of the Intermediate Value Theorem are satisfied. So, there exists a number c in $(0,8)$ such that $f(c)=1$. This means that c satisfies $\sqrt[3]{c}+c=1$, in otherwords, is a solution for the equation given. ♣

Example 3.56: Roots of Function

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

Solution. By theorem 3.8, f is continuous. Since $f(0)=-2$ and $f(1)=3$, and 0 is between -2 and 3, there is a $c\in(0,1)$ such that $f(c)=0$. ♣

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

Example 3.57: Approximating Roots

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

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

Exercises for 3.7

Exercise 3.7.1 *Consider the function*

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

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

Exercise 3.7.2 *Find the values of a that make the function $f(x)$ continuous for all real numbers.*

$$f(x)=\begin{cases} 4x+5, & \text{if } x\geq -2,\\ x^2+a, & \text{if } x<-2.\end{cases}$$

Exercise 3.7.3 *Find the values of the constant c so that the function* $g(x)$ *is continuous on* $(-\infty,\infty)$*, where*

$$g(x) = \begin{cases} 2-2c^2x, & \text{if } x < -1, \\ 6-7cx^2, & \text{if } x \geq -1. \end{cases}$$

Exercise 3.7.4 *Approximate a root of* $f = x^3 - 4x^2 + 2x + 2$ *to one decimal place.*

Exercise 3.7.5 *Approximate a root of* $f = x^4 + x^3 - 5x + 1$ *to one decimal place.*

Exercise 3.7.6 *Show that the equation* $\sqrt[3]{x} + x = 1$ *has a solution in the interval* $(0,8)$.

4. Derivatives

4.1 The Rate of Change of a Function

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

Example 4.1: Small Changes in x

Consider $y = f(x) = \sqrt{625 - x^2}$ (the upper semicircle of radius 25 centered at the origin), and let's compute the changes of y resulting from small changes of x around $x = 7$.

Solution. When $x = 7$, we find that $y = \sqrt{625 - 49} = 24$. Suppose we want to know how much y changes when x increases a little, say to 7.1 or 7.01.

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

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

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

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

The quantity $\Delta y/\Delta x \approx -0.294$ may be interpreted as the slope of the line through $(7, 24)$ and $(7.1, 23.9706)$, called a **chord** of the circle. In general, if we draw the chord from the point $(7, 24)$ to a nearby point on the semicircle $(7 + \Delta x, f(7 + \Delta x))$, the slope of this chord is the so-called **difference quotient**

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

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

As Δx is made smaller (closer to 0), $7+\Delta x$ gets closer to 7 and the chord joining $(7,f(7))$ to $(7+\Delta x, f(7+\Delta x))$ shifts slightly, as shown in Figure 4.1. The chord gets closer and closer to the **tangent line** to the circle at the point $(7,24)$. (The tangent line is the line that just grazes the circle at that point, i.e., it doesn't meet the circle at any second point.) Thus, as Δx gets smaller and smaller, the slope $\Delta y/\Delta x$ of the chord gets closer and closer to the slope of the tangent line. This is actually quite difficult to see when Δx is small, because of the scale of the graph. The values of Δx used for the figure are 1, 5, 10 and 15, not really very small values. The tangent line is the one that is uppermost at the right hand endpoint.

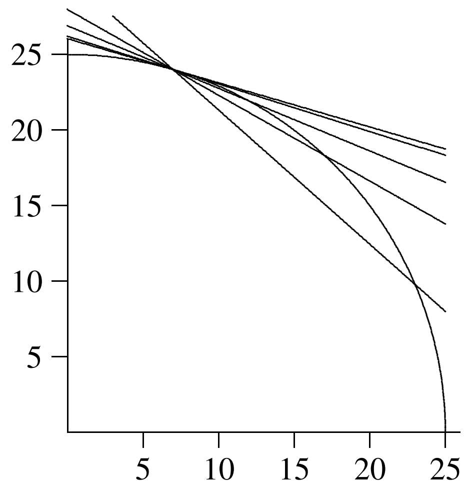


Figure 4.1: Chords approximating the tangent line.

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

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

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

$$= \frac{-14-\Delta x}{\sqrt{625-(7+\Delta x)^2}+24}$$

Now, can we tell by looking at this last formula what happens when Δx gets very close to zero? The numerator clearly gets very close to -14 while the denominator gets very close to $\sqrt{625-7^2}+24 = 48$. The fraction is therefore very close to $-14/48 = -7/24 \cong -0.29167$. In fact, the slope of the tangent line is exactly $-7/24$.

What about the slope of the tangent line at $x = 12$? Well, 12 can't be all that different from 7; we just have to redo the calculation with 12 instead of 7. This won't be hard, but it will be a bit tedious. What if we try to do all the algebra without using a specific value for x? Let's copy from above, replacing 7 by x.

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

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

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

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

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

The expression $-x/\sqrt{625-x^2}$ defines a new function called the **derivative** of the original function (since it is derived from the original function). If the original is referred to as f or y then the derivative is often written f' or y' (pronounced "f prime" or "y prime"). So in this case we might write $f'(x) = -x/\sqrt{625-x^2}$ or $y' = -x/\sqrt{625-x^2}$. At a particular point, say $x = 7$, we write $f'(7) = -7/24$ and we say that "f prime of 7 is $-7/24$" or "the derivative of f at 7 is $-7/24$."

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

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

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

At this point, we should note that the idea of letting Δx get closer and closer to 0 is precisely the idea of a limit that we discussed in the last chapter. The limit here is a limit as Δx approaches 0. Using limit notation, we can write $f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x}$.

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

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

Velocity

We started this section by saying, "It is often useful to know how sensitive the value of y is to small changes in x." We have seen one purely mathematical example of this, involving the function $f(x) = \sqrt{625-x^2}$. Here is a more applied example.

With careful measurement it might be possible to discover that the height of a dropped ball t seconds after it is released is $h(t) = h_0 - kt^2$. (Here h_0 is the initial height of the ball, when $t = 0$, and k is some number determined by the experiment.) A natural question is then, "How fast is the ball going at time t?" We can certainly get a pretty good idea with a little simple arithmetic. To make the calculation more concrete, let's use units of meters and seconds and say that $h_0 = 100$ meters and $k = 4.9$. Suppose we're interested in the speed at $t = 2$. We know that when $t = 2$ the height is $100 - 4 \cdot 4.9 = 80.4$ meters. A second later, at $t = 3$, the height is $100 - 9 \cdot 4.9 = 55.9$ meters. The change in height during that second is $55.9 - 80.4 = -24.5$ meters. The negative sign means the height has decreased, as we expect for a falling ball, and the number 24.5 is the average speed of the ball during the time interval, in meters per second.

We might guess that 24.5 meters per second is not a terrible estimate of the speed at $t = 2$, but certainly we can do better. At $t = 2.5$ the height is $100 - 4.9(2.5)^2 = 69.375$ meters. During the half second from $t = 2$ to $t = 2.5$, the change in height is $69.375 - 80.4 = -11.025$ meters giving an average speed of $11.025/(1/2) = 22.05$ meters per second. This should be a better estimate of the speed at $t = 2$. So it's clear now how to get better and better approximations: compute average speeds over shorter and shorter time intervals. Between $t = 2$ and $t = 2.01$, for example, the ball drops 0.19649 meters in one hundredth of a second, at an average speed of 19.649 meters per second.

We still might reasonably ask for the precise speed at $t = 2$ (the *instantaneous* speed) rather than just an approximation to it. For this, once again, we need a limit. Let's calculate the average speed during the time interval from $t = 2$ to $t = 2 + \Delta t$ without specifying a particular value for Δt. The change in height during the time interval from $t = 2$ to $t = 2 + \Delta t$ is

$$\begin{aligned} h(2+\Delta t) - h(2) &= (100 - 4.9(2+\Delta t)^2) - 80.4 \\ &= 100 - 4.9(4 + 4\Delta t + \Delta t^2) - 80.4 \end{aligned}$$

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

The average speed during this time interval is then

$$\frac{\Delta t(19.6 + 4.9\Delta t)}{\Delta t} = 19.6 + 4.9\Delta t.$$

When Δt is very small, this is very close to 19.6. Indeed, $\lim_{\Delta x \to 0}(19.6 + 4.9\Delta t) = 19.6$. So the exact speed at $t = 2$ is 19.6 meters per second.

At this stage we need to make a distinction between *speed* and *velocity*. Velocity is signed speed, that is, speed with a direction indicated by a sign (positive or negative). Our algebra above actually told us that the instantaneous velocity of the ball at $t = 2$ is -19.6 meters per second. The number 19.6 is the speed and the negative sign indicates that the motion is directed downwards (the direction of decreasing height).

In the language of the previous section, we might have started with $f(x) = 100 - 4.9x^2$ and asked for the slope of the tangent line at $x = 2$. We would have answered that question by computing

$$\lim_{\Delta x \to 0} \frac{f(2 + \Delta x) - f(2)}{\Delta x} = \lim_{\Delta x \to 0} \frac{-19.6\Delta x - 4.9\Delta x^2}{\Delta x} = -19.6 - 4.9\Delta x = 19.6$$

The algebra is the same. Thus, the velocity of the ball is the value of the derivative of a certain function, namely, of the function that gives the position of the ball.

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

Exercises for Section 4.1

Exercise 4.1.1 *Draw the graph of the function $y = f(x) = \sqrt{169 - x^2}$ between $x = 0$ and $x = 13$. Find the slope $\Delta y/\Delta x$ of the chord between the points of the circle lying over (a) $x = 12$ and $x = 13$, (b) $x = 12$ and $x = 12.1$, (c) $x = 12$ and $x = 12.01$, (d) $x = 12$ and $x = 12.001$. Now use the geometry of tangent lines on a circle to find (e) the exact value of the derivative $f'(12)$. Your answers to (a)–(d) should be getting closer and closer to your answer to (e).*

Exercise 4.1.2 *Use geometry to find the derivative $f'(x)$ of the function $f(x) = \sqrt{625 - x^2}$ in the text for each of the following x: (a) 20, (b) 24, (c) -7, (d) -15. Draw a graph of the upper semicircle, and draw the tangent line at each of these four points.*

Exercise 4.1.3 *Draw the graph of the function $y = f(x) = 1/x$ between $x = 1/2$ and $x = 4$. Find the slope of the chord between (a) $x = 3$ and $x = 3.1$, (b) $x = 3$ and $x = 3.01$, (c) $x = 3$ and $x = 3.001$. Now* *use algebra to find a simple formula for the slope of the chord between* $(3, f(3))$ *and* $(3+\Delta x, f(3+\Delta x))$*. Determine what happens when* Δx *approaches 0. In your graph of* $y = 1/x$*, draw the straight line through the point* $(3, 1/3)$ *whose slope is this limiting value of the difference quotient as* Δx *approaches 0.*

Exercise 4.1.4 *Find an algebraic expression for the difference quotient* $\big(f(1+\Delta x) - f(1)\big)/\Delta x$ *when* $f(x) = x^2 - (1/x)$*. Simplify the expression as much as possible. Then determine what happens as* Δx *approaches 0. That value is* $f'(1)$*.*

Exercise 4.1.5 *Draw the graph of* $y = f(x) = x^3$ *between* $x = 0$ *and* $x = 1.5$*. Find the slope of the chord between (a)* $x = 1$ *and* $x = 1.1$*, (b)* $x = 1$ *and* $x = 1.001$*, (c)* $x = 1$ *and* $x = 1.00001$*. Then use algebra to find a simple formula for the slope of the chord between* 1 *and* $1+\Delta x$*. (Use the expansion* $(A+B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$*.) Determine what happens as* Δx *approaches 0, and in your graph of* $y = x^3$ *draw the straight line through the point* $(1,1)$ *whose slope is equal to the value you just found.*

Exercise 4.1.6 *Find an algebraic expression for the difference quotient* $(f(x+\Delta x) - f(x))/\Delta x$ *when* $f(x) = mx + b$*. Simplify the expression as much as possible. Then determine what happens as* Δx *approaches 0. That value is* $f'(x)$*.*

Exercise 4.1.7 *Sketch the unit circle. Discuss the behavior of the slope of the tangent line at various angles around the circle. Which trigonometric function gives the slope of the tangent line at an angle* θ*? Why? Hint: think in terms of ratios of sides of triangles.*

Exercise 4.1.8 *Sketch the parabola* $y = x^2$*. For what values of* x *on the parabola is the slope of the tangent line positive? Negative? What do you notice about the graph at the point(s) where the sign of the slope changes from positive to negative and vice versa?*

Exercise 4.1.9 *An object is traveling in a straight line so that its position (that is, distance from some fixed point) is given by this table:*

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

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

Exercise 4.1.10 *Let* $y = f(t) = t^2$*, where* t *is the time in seconds and* y *is the distance in meters that an object falls on a certain airless planet. Draw a graph of this function between* $t = 0$ *and* $t = 3$*. Make a table of the average speed of the falling object between (a) 2 sec and 3 sec, (b) 2 sec and 2.1 sec, (c) 2 sec and 2.01 sec, (d) 2 sec and 2.001 sec. Then use algebra to find a simple formula for the average speed between time* 2 *and time* $2+\Delta t$*. (If you substitute* $\Delta t = 1, 0.1, 0.01, 0.001$ *in this formula you should again get the answers to parts (a)–(d).) Next, in your formula for average speed (which should be in simplified form) determine what happens as* Δt *approaches zero. This is the instantaneous speed. Finally, in your graph of* $y = t^2$ *draw the straight line through the point* $(2,4)$ *whose slope is the instantaneous velocity you just computed; it should of course be the tangent line.*

Exercise 4.1.11 *If an object is dropped from an 80-meter high window, its height* y *above the ground at time* t *seconds is given by the formula* $y = f(t) = 80 - 4.9t^2$*. (Here we are neglecting air resistance; the* *graph of this function was shown in figure 1.1.) Find the average velocity of the falling object between (a) 1 sec and 1.1 sec, (b) 1 sec and 1.01 sec, (c) 1 sec and 1.001 sec. Now use algebra to find a simple formula for the average velocity of the falling object between 1 sec and $1+\Delta t$ sec. Determine what happens to this average velocity as Δt approaches 0. That is the instantaneous velocity at time $t = 1$ second (it will be negative, because the object is falling).*

4.2 The Derivative Function

In Section 4.1, we have seen how to create, or derive, a new function $f'(x)$ from a function $f(x)$, and that this new function carries important information. In one example we saw that $f'(x)$ tells us how steep the graph of $f(x)$ is; in another we saw that $f'(x)$ tells us the velocity of an object if $f(x)$ tells us the position of the object at time x. As we said earlier, this same mathematical idea is useful whenever $f(x)$ represents some changing quantity and we want to know something about how it changes, or roughly, the "rate" at which it changes. Most functions encountered in practice are built up from a small collection of "primitive" functions in a few simple ways, for example, by adding or multiplying functions together to get new, more complicated functions. To make good use of the information provided by $f'(x)$ we need to be able to compute it for a variety of such functions.

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

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

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

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

Definition 4.2: Definition of Derivative

The derivative of $y = f(x)$ with respect to x is

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

Some textbooks use h in place of Δx in the definition of derivative:

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

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

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

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

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

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

Example 4.3: Derivative of $y = t^2$

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

Solution. We compute

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

♣

Remember that Δt is a single quantity, not a "Δ" times a "t", and so Δt^2 is $(\Delta t)^2$ not $\Delta(t^2)$. Doing the same example using the second formula for the derivative with h in place of Δt gives the following. Note that we compute $f(t+h)$ by substituting $t+h$ in place of t everywhere we see t in the expression $f(t)$, *while making no other changes* (at least initially). For example, if $f(t) = t + \sqrt{(t+3)^2 - t}$ then $f(t+h) = (t+h) + \sqrt{((t+h)+3)^2 - (t+h)} = t + h + \sqrt{(t+h+3)^2 - t - h}$.

Example 4.4: Derivative of $y = t^2$

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

Solution. We compute

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

♣

Example 4.5: Derivative

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

Solution. The computation:

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

♣

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

To recap, given any function f and any number x in the domain of f, we define $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ wherever this limit exists, and we call the number $f'(x)$ the derivative of f at x. Geometrically, $f'(x)$ is the slope of the tangent line to the graph of f at the point $(x, f(x))$. The following symbols also represent the derivative:

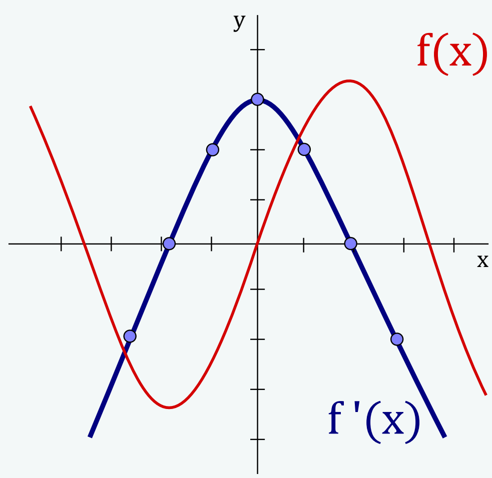
$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x).$$

The symbol d/dx is called a differential operator which means to take the derivative of the function $f(x)$ with respect to the variable x.

In the next example we emphasize the geometrical interpretation of derivative.

Example 4.6: Geometrical Interpretation of Derivative

Consider the function $f(x)$ given by the graph below. Verify that the graph of $f'(x)$ is indeed the derivative of $f(x)$ by analyzing slopes of tangent lines to the graph at different points.



<u>**Solution.**</u> We must think about the tangent lines to the graph of f, because the slopes of these lines are the values of $f'(x)$.

We start by checking the graph of f for horizontal tangent lines, since horizontal lines have a slope of 0. We find that the tangent line is horizontal at the points where x has the values -1.9 and 1.8 (approximately). At each of these values of x, we must have $f'(x) = 0$, which means that the graph of f' has an x-intercept (a point where the graph intersects the x-axis).

Note that horizontal tangent lines have a slope of zero and these occur approximately at the points $(-1.9, -3.2)$ and $(1.8, 3.2)$ of the graph. Therefore $f'(x)$ will cross the x-axis when $x = -1.9$ and $x = 1.8$.

Analyzing the slope of the tangent line of $f(x)$ at $x = 0$ gives approximately 3.0, thus, $f'(0) = 3.0$. Similarly, analyzing the slope of the tangent lines of $f(x)$ at $x = 1$ and $x = -1$ give approximately 2.0 for both, thus, $f'(1) = f'(-1) = 2.0$. ♣

In the next example we verify that the slope of a straight line is m.

Example 4.7: Derivative of a Linear Function

Let m, b be any two real numbers. Determine $f'(x)$ if $f(x) = mx + b$.

Solution. By the definition of derivative (using h in place of Δx) we have,

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

$$= \lim_{h\to 0} \frac{mh}{h} = \lim_{h\to 0} m = m.$$

This is not surprising. We know that $f'(x)$ always represents the slope of a tangent line to the graph of f. In this example, since the graph of f is a straight line $y = mx + b$ already, every tangent line is the same line $y = mx + b$. Since this line has a slope of m, we must have $f'(x) = m$. ♣

4.2.1. Differentiable

Now that we have introduced the derivative of a function at a point, we can begin to use the adjective **differentiable**.

Definition 4.8: Differentiable at a Point

A function f is differentiable at point a if $f'(a)$ exists.

Definition 4.9: Differentiable on an Interval

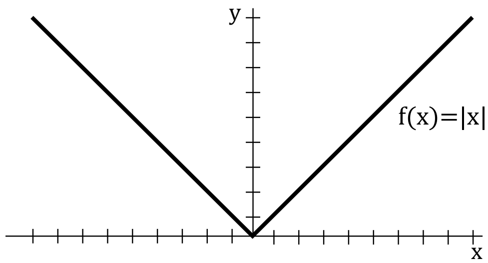
A function f is differentiable on an open interval if it is differentiable at every point in the interval.

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

Example 4.10: Derivative of the Absolute Value

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

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



We can summarize this as

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

In particular, the absolute value function $f(x) = |x|$ is *not* differentiable at $x = 0$. ♣

We note that the following theorem can be proved using limits.

Theorem 4.11: Differentiable implies Continuity

If f is differentiable at a, then f is continuous at a.

Proof. Suppose that f is differentiable at a. That is,

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

exists. At this stage, we find it convenient to write this limit in an alternative form so that its connection with continuity can become more easily seen. If we let $x = a + h$, then $h = x - a$. Furthermore, $h \to 0$ is equivalent to $x \to a$. So,

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

(This alternative formulation of the derivative is also standard. We will use it whenever we find it convenient to do so. You should get familiar with it.) Continuity at a can now be proved as follows:

$$\begin{aligned} \lim_{x \to a} f(x) &= \lim_{x \to a} \left(\frac{f(x) - f(a)}{x - a} \cdot (x - a) + f(a) \right) \\ &= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \cdot \lim_{x \to a} (x - a) + \lim_{x \to a} f(a) \\ &= f'(a) \cdot (a - a) + f(a) \\ &= f(a). \end{aligned}$$

♣

However, if f is continuous at a it is *not* necessarily true that f is differentiable at a. For example, it was shown that $f(x) = |x|$ is not differentiable at $x = 0$ in the previous example, however, one can observe that $f(x) = |x|$ is continuous everywhere.

Example 4.12: Derivative of $y = x^{2/3}$

Discuss the derivative of the function $y = x^{2/3}$, shown in Figure 4.2.

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

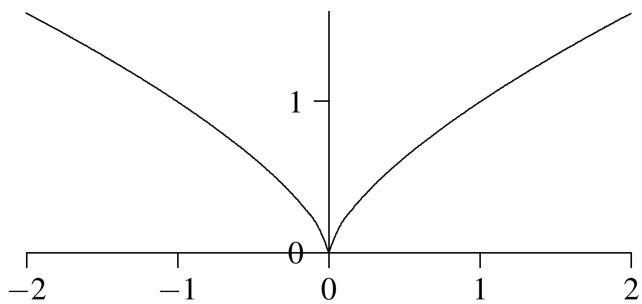


Figure 4.2: A cusp on $x^{2/3}$.

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

4.2.2. Second and Other Derivatives

If f is a differentiable function then its derivative f' is also a function and so we can take the derivative of f'. The new function, denoted by f'', is called the **second derivative** of f, since it is the derivative of the derivative of f.

The following symbols represent the second derivative:

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

We can continue this process to get the third derivative of f.

In general, the **nth derivative** of f is denoted by $f^{(n)}$ and is obtained from f by differentiating n times. If $y = f(x)$, then we write:

$$y^{(n)} = f^{(n)}(x) = \frac{d^ny}{dx^n}.$$

4.2.3. Velocities

Suppose $f(t)$ is a position function of an object, representing the displacement of the object from the origin at time t. In terms of derivatives, the **velocity of an object is:**

$$v(a) = f'(a)$$

The change of velocity with respect to time is called the **acceleration** and can be found as follows:

$$a(t) = v'(t) = f''(t).$$

Acceleration is the derivative of the velocity function and the second derivative of the position function.

Example 4.13: Position, Velocity and Acceleration

Suppose the position function of an object is $f(t) = t^2$ metres at t seconds. Find the velocity and acceleration of the object at time $t = 1s$.

Solution. By the definition of velocity and acceleration we need to compute $f'(t)$ and $f''(t)$. Using the definition of derivative, we have,

$$f'(t) = \lim_{h\to 0}\frac{(t+h)^2 - t^2}{h} = \lim_{h\to 0}\frac{2th+h^2}{h} = \lim_{h\to 0}(2t+h) = 2t.$$

Therefore, $v(t) = f'(t) = 2t$. Thus, the velocity at time $t = 1$ is $v(1) = 2\ m/s$. We now have that the acceleration at time t is:

$$a(t) = f''(t) = \lim_{h\to 0}\frac{2(t+h) - 2t}{h} = \lim_{h\to 0}\frac{2h}{h} = 2.$$

Therefore, $a(t) = 2$. Substituting $t = 1$ into the function $a(t)$ gives $a(1) = 2\ m/s^2$. ♣

Exercises for Section 4.2

Exercise 4.2.1 *Find the derivatives of the following functions.*

(a) $y = f(x) = \sqrt{169 - x^2}$

(b) $y = f(t) = 80 - 4.9t^2$

(c) $y = f(x) = x^2 - (1/x)$

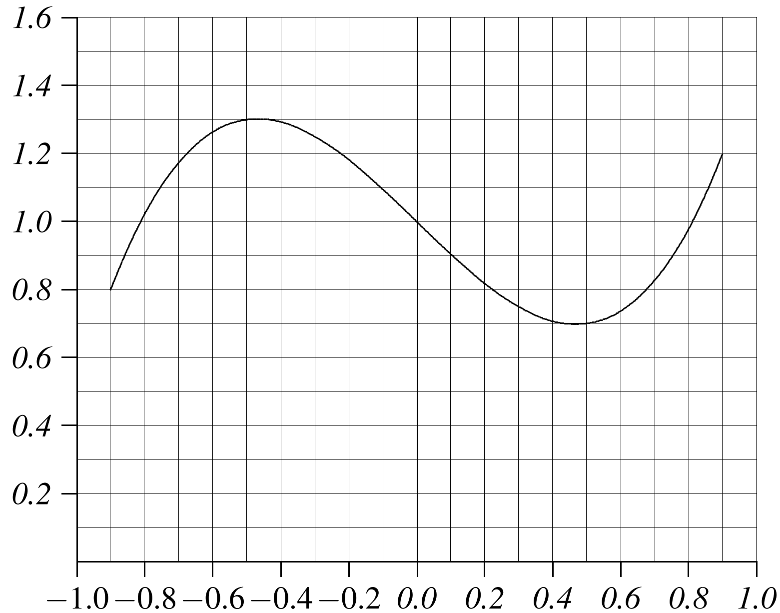
(d) $y = f(x) = ax^2 + bx + c$, *where a, b, and c are constants.*

(e) $y = f(x) = x^3$

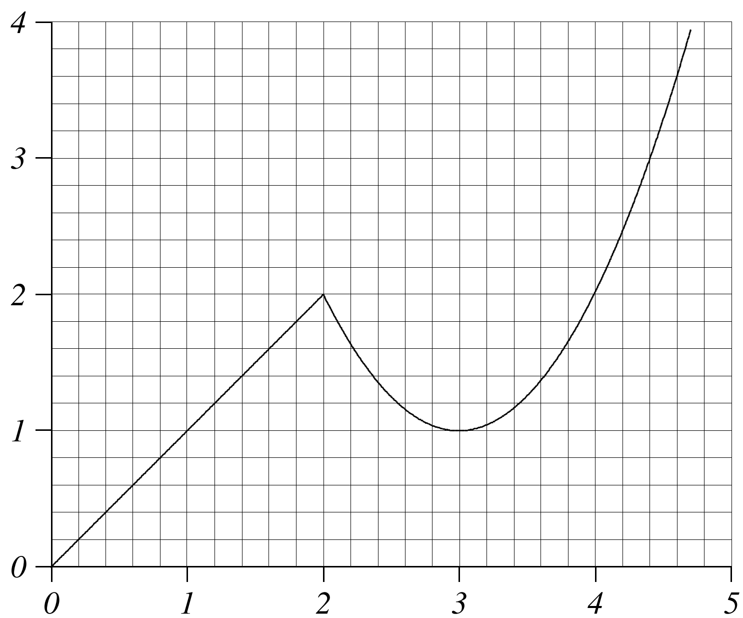
(f) $y = f(x) = 2/\sqrt{2x+1}$

(g) $y = g(t) = (2t-1)/(t+2)$

Exercise 4.2.2 *Shown is the graph of a function $f(x)$. Sketch the graph of $f'(x)$ by estimating the derivative at a number of points in the interval: estimate the derivative at regular intervals from one end of the interval to the other, and also at "special" points, as when the derivative is zero. Make sure you indicate any places where the derivative does not exist.*



Exercise 4.2.3 *Shown is the graph of a function $f(x)$. Sketch the graph of $f'(x)$ by estimating the derivative at a number of points in the interval: estimate the derivative at regular intervals from one end of the interval to the other, and also at "special" points, as when the derivative is zero. Make sure you indicate any places where the derivative does not exist.*



Exercise 4.2.4 *Find an equation for the tangent line to the graph of* $f(x) = 5 - x - 3x^2$ *at the point* $x = 2$

Exercise 4.2.5 *Find a value for a so that the graph of* $f(x) = x^2 + ax - 3$ *has a horizontal tangent line at* $x = 4$.

4.3 Derivative Rules

Using the definition of the derivative of a function is quite tedious. In this section we introduce a number of different shortcuts that can be used to compute the derivative. Recall that the *definition of derivative* is:

Given any number x for which the limit

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

exists, we assign to x the number $f'(x)$.

Next, we give some basic *derivative rules* for finding derivatives without having to use the limit definition directly.

Theorem 4.14: Derivative of a Constant Function

Let c *be a constant, then* $\dfrac{d}{dx}(c) = 0$.

Proof. Let $f(x) = c$ be a constant function. By the definition of derivative:

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

♣

Example 4.15: Derivative of a Constant Function

The derivative of $f(x) = 17$ *is* $f'(x) = 0$ *since the derivative of a constant is* 0.

Theorem 4.16: The Power Rule

If n *is a positive integer, then* $\dfrac{d}{dx}(x^n) = nx^{n-1}$.

Proof. We use the formula:

$$x^n - a^n = (x-a)(x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1})$$

which can be verified by mulitplying out the right side. Let $f(x) = x^n$ be a power function for some positive integer n. Then at any number a we have:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{x^n - a^n}{x - a} = \lim_{x \to a}(x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}) = na^{n-1}.$$

♣

It turns out that the Power Rule holds for any real number n (though it is a bit more difficult to prove).

Theorem 4.17: The Power Rule (General)

If n is any real number, then $\frac{d}{dx}(x^n) = nx^{n-1}$.

Example 4.18: Derivative of a Power Function

By the power rule, the derivative of $g(x) = x^4$ is $g'(x) = 4x^3$.

Theorem 4.19: The Constant Multiple Rule

If c is a constant and f is a differentiable function, then

$$\frac{d}{dx}[cf(x)] = c\frac{d}{dx}f(x).$$

Proof. For convenience let $g(x) = cf(x)$. Then:

$$g'(x) = \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = \lim_{h \to 0} \frac{cf(x+h) - cf(x)}{h}$$
$$= \lim_{h \to 0} c\left[\frac{f(x+h) - f(x)}{h}\right] = c\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = cf'(x),$$

where c can be moved in front of the limit by the Limit Rules. ♣

Example 4.20: Derivative of a Multiple of a Function

By the constant multiple rule and the previous example, the derivative of $F(x) = 2 \cdot (17 + x^4)$ is

$$F'(x) = 2(4x^3) = 8x^3.$$

Theorem 4.21: The Sum/Difference Rule

If f and g are both differentiable functions, then

$$\frac{d}{dx}[f(x) \pm g(x)] = \frac{d}{dx}f(x) \pm \frac{d}{dx}g(x).$$

Proof. For convenience let $r(x) = f(x) \pm g(x)$. Then:

$$\begin{aligned} r'(x) &= \lim_{h\to 0} \frac{r(x+h)-r(x)}{h} \\ &= \lim_{h\to 0} \frac{[f(x+h) \pm g(x+h)] - [f(x) \pm g(x)]}{h} \\ &= \lim_{h\to 0} \left[\frac{f(x+h)-f(x)}{h} \pm \frac{g(x+h)-g(x)}{h} \right] \\ &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \pm \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} \\ &= f'(x) + g'(x) \end{aligned}$$

♣

Example 4.22: Derivative of a Sum/Difference of Functions

By the sum/difference rule, the derivative of $h(x) = 17 + x^4$ is

$$h'(x) = f'(x) + g'(x) = 0 + 4x^3 = 4x^3.$$

Theorem 4.23: The Product Rule

If f and g are both differentiable functions, then

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

Proof. For convenience let $r(x) = f(x) \cdot g(x)$. As in the previous proof, we want to separate the functions f and g. The trick is to add and subtract $f(x+h)g(x)$ in the numerator. Then:

$$\begin{aligned} r'(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \\ &= \lim_{h\to 0} \left[f(x+h)\frac{g(x+h)-g(x)}{h} + g(x)\frac{f(x+h)-f(x)}{h} \right] \\ &= \lim_{h\to 0} f(x+h) \cdot \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} + \lim_{h\to 0} g(x) \cdot \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= f(x)g'(x) + g(x)f'(x) \end{aligned}$$

♣

Example 4.24: Derivative of a Product of Functions

Find the derivative of $h(x) = (3x-1)(2x+3)$.

Solution. One way to do this question is to expand the expression. Alternatively, we use the product rule with $f(x) = 3x-1$ and $g(x) = 2x+3$. Note that $f'(x) = 3$ and $g'(x) = 2$, so,

$$h'(x) = (3)\cdot(2x+3) + (3x-1)\cdot(2) = 6x+9+6x-2 = 12x+7.$$

♣

Theorem 4.25: The Quotient Rule

If f and g are both differentiable functions, then

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

Proof. The proof is similar to the previous proof but the trick is to add and subtract the term $f(x)g(x)$ in the numerator. We omit the details. ♣

Example 4.26: Derivative of a Quotient of Functions

Find the derivative of $h(x) = \dfrac{3x-1}{2x+3}$.

Solution. By the quotient rule (using $f(x) = 3x-1$ and $g(x) = 2x+3$) we have:

$$\begin{aligned} h'(x) &= \frac{\frac{d}{dx}(3x-1)\cdot(2x+3) - (3x-1)\cdot\frac{d}{dx}(2x+3)}{(2x+3)^2} \\ &= \frac{3(2x+3)-(3x-1)(2)}{(2x+3)^2} = \frac{11}{(2x+3)^2}. \end{aligned}$$

♣

Example 4.27: Second Derivative

Find the second derivative of $f(x) = 5x^3 + 3x^2$.

Solution. We must differentiate $f(x)$ twice:

$$f'(x) = 15x^2 + 6x,$$

$$f''(x) = 30x + 6.$$

♣

Exercises for Section 4.3

Exercise 4.3.1 *Find the derivatives of the following functions.*

(a) x^{100}

(b) x^{-100}

(c) $\dfrac{1}{x^5}$

(d) x^{π}

(e) $x^{3/4}$

(f) $x^{-9/7}$

(g) $5x^3+12x^2-15$

(h) $-4x^5+3x^2-5/x^2$

(i) $5(-3x^2+5x+1)$

(j) $(x+1)(x^2+2x-3)$

(k) $(x+1)(x^2+2x-3)^{-1}$

(l) $x^3(x^3-5x+10)$

(m) $(x^2+5x-3)(x^5)$

(n) $(x^2+5x-3)(x^{-5})$

(o) $(5x^3+12x^2-15)^{-1}$

Exercise 4.3.2 *Find an equation for the tangent line to $f(x)=x^3/4-1/x$ at $x=-2$.*

Exercise 4.3.3 *Find an equation for the tangent line to $f(x)=3x^2-\pi^3$ at $x=4$.*

Exercise 4.3.4 *Suppose the position of an object at time t is given by $f(t)=-49t^2/10+5t+10$. Find a function giving the speed of the object at time t. The acceleration of an object is the rate at which its speed is changing, which means it is given by the derivative of the speed function. Find the acceleration of the object at time t.*

Exercise 4.3.5 *Let $f(x)=x^3$ and $c=3$. Sketch the graphs of f, cf, f', and $(cf)'$ on the same diagram.*

Exercise 4.3.6 *The general polynomial P of degree n in the variable x has the form $P(x)=\sum_{k=0}^{n} a_k x^k = a_0+a_1x+\ldots+a_nx^n$. What is the derivative (with respect to x) of P?*

Exercise 4.3.7 *Find a cubic polynomial whose graph has horizontal tangents at $(-2,5)$ and $(2,3)$.*

Exercise 4.3.8 *Prove that $\dfrac{d}{dx}(cf(x))=cf'(x)$ using the definition of the derivative.*

Exercise 4.3.9 *Suppose that f and g are differentiable at x. Show that $f-g$ is differentiable at x using the two linearity properties from this section.*

Exercise 4.3.10 *Use the product rule to compute the derivative of $f(x)=(2x-3)^2$. Sketch the function. Find an equation of the tangent line to the curve at $x=2$. Sketch the tangent line at $x=2$.*

Exercise 4.3.11 *Suppose that f, g, and h are differentiable functions. Show that $(fgh)'(x)=f'(x)g(x)h(x)+f(x)g'(x)h(x)+f(x)g(x)h'(x)$.*

Exercise 4.3.12 *Compute the derivative of $\dfrac{x^3}{x^3-5x+10}$.*

Exercise 4.3.13 *Compute the derivative of* $\dfrac{x^2+5x-3}{x^5-6x^3+3x^2-7x+1}$.

Exercise 4.3.14 *Compute the derivative of* $\dfrac{x}{\sqrt{x-625}}$.

Exercise 4.3.15 *Compute the derivative of* $\dfrac{\sqrt{x-5}}{x^{20}}$.

Exercise 4.3.16 *Find an equation for the tangent line to* $f(x)=(x^2-4)/(5-x)$ *at* $x=3$.

Exercise 4.3.17 *Find an equation for the tangent line to* $f(x)=(x-2)/(x^3+4x-1)$ *at* $x=1$.

Exercise 4.3.18 *If* $f'(4)=5$, $g'(4)=12$, $(fg)(4)=f(4)g(4)=2$, *and* $g(4)=6$, *compute* $f(4)$ *and* $\dfrac{d}{dx}\dfrac{f}{g}$ *at 4.*

4.4 Derivative Rules for Trigonometric Functions

We next look at the derivative of the sine function. In order to prove the derivative formula for sine, we recall two limit computations from earlier:

$$\lim_{x\to 0}\frac{\sin x}{x}=1 \qquad \text{and} \qquad \lim_{x\to 0}\frac{\cos x-1}{x}=0,$$

and the double angle formula

$$\sin(A+B)=\sin A\cos B+\sin B\cos A.$$

Theorem 4.28: Derivative of Sine Function

$(\sin x)'=\cos x$

Proof. Let $f(x) = \sin x$. Using the definition of derivative we have:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x+h)-\sin x}{h} \\ &= \lim_{h\to 0} \frac{\sin x\cos h + \cos x \sin h - \sin x}{h} \\ &= \lim_{h\to 0} \sin x \cdot \lim_{h\to 0} \frac{\cos h - 1}{h} + \lim_{h\to 0} \cos x \cdot \lim_{h\to 0} \frac{\sin h}{h} \\ &= \sin x \cdot 0 + \cos x \cdot 1 \\ &= \cos x \end{aligned}$$

since

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

♣

A formula for the derivative of the *cosine function* can be found in a similar fashion:

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

Using the quotient rule we get formulas for the remaining trigonometric ratios. To summarize, here are the derivatives of the six trigonometric functions:

$\frac{d}{dx}(\sin(x)) = \cos(x)$	$\frac{d}{dx}(\tan(x)) = \sec^2(x)$	$\frac{d}{dx}(\csc(x)) = -\csc(x)\cot(x)$
$\frac{d}{dx}(\cos(x)) = -\sin(x)$	$\frac{d}{dx}(\cot(x)) = -\csc^2(x)$	$\frac{d}{dx}(\sec(x)) = \sec(x)\tan(x)$

Example 4.29: Derivative of Product of Trigonometric Functions

Find the derivative of $f(x) = \sin x \tan x$.

Solution. Using the Product Rule we obtain

$$f'(x) = \cos x \tan x + \sin x \sec^2 x.$$

♣

Exercises for Section 4.4

Exercise 4.4.1 *Find the derivatives of the following functions.*

(a) $\sin x \cos x$ *(b)* $\cot x$ *(c)* $\csc x - x \tan x$

Exercise 4.4.2 *Find the points on the curve* $y = x + 2\cos x$ *that have a horizontal tangent line.*

4.5 The Chain Rule

Let $h(x) = \sqrt{625 - x^2}$. The rules stated previously do not allow us to find $h'(x)$. However, $h(x)$ is a composition of two functions. Let $f(x) = \sqrt{x}$ and $g(x) = 625 - x^2$. Then we see that

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

From our rules we know that $f'(x) = \frac{1}{2}x^{-1/2}$ and $g'(x) = -2x$, thus it would be convenient to have a rule which allows us to differentiate $f \circ g$ in terms of f' and g'. This gives rise to the chain rule.

The Chain Rule

If g is differentiable at x and f is differentiable at $g(x)$, then the composite function $h = f \circ g$ [recall $f \circ g$ is defined as $f(g(x))$] is differentiable at x and $h'(x)$ is given by:

$$h'(x) = f'(g(x)) \cdot g'(x).$$

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

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

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

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

Example 4.30: Chain Rule

Compute the derivative of $\sqrt{625 - x^2}$.

Solution. We already know that the answer is $-x/\sqrt{625-x^2}$, computed directly from the limit. In the context of the chain rule, we have $f(x)=\sqrt{x}$, $g(x)=625-x^2$. We know that $f'(x)=(1/2)x^{-1/2}$, so $f'(g(x))=(1/2)(625-x^2)^{-1/2}$. Note that this is a two step computation: first compute $f'(x)$, then replace x by $g(x)$. Since $g'(x)=-2x$ we have

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

♣

Example 4.31: Chain Rule

Compute the derivative of $1/\sqrt{625-x^2}$.

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

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

♣

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

Example 4.32: Derivative of Quotient

Compute the derivative of

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

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

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

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

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

Finally, we use the chain rule:

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

And putting it all together:

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

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

Example 4.33: Chain of Composition

Compute the derivative of $\sqrt{1+\sqrt{1+\sqrt{x}}}$.

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

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

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

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

So the original derivative is

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

♣

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

Example 4.34: Derivative of Quotient without Quotient Rule

Compute the derivative of $f(x)=\dfrac{x^3}{x^2+1}$.

Solution. Write $f(x) = x^3(x^2+1)^{-1}$, then

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

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

♣

Exercises for Section 4.5

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

Exercise 4.5.1 $x^4 - 3x^3 + (1/2)x^2 + 7x - \pi$

Exercise 4.5.2 $x^3 - 2x^2 + 4\sqrt{x}$

Exercise 4.5.3 $(x^2+1)^3$

Exercise 4.5.4 $x\sqrt{169-x^2}$

Exercise 4.5.5 $(x^2-4x+5)\sqrt{25-x^2}$

Exercise 4.5.6 $\sqrt{r^2-x^2}$, *r is a constant*

Exercise 4.5.7 $\sqrt{1+x^4}$

Exercise 4.5.8 $\dfrac{1}{\sqrt{5-\sqrt{x}}}$.

Exercise 4.5.9 $(1+3x)^2$

Exercise 4.5.10 $\dfrac{(x^2+x+1)}{(1-x)}$

Exercise 4.5.11 $\dfrac{\sqrt{25-x^2}}{x}$

Exercise 4.5.12 $\sqrt{\dfrac{169}{x}-x}$

Exercise 4.5.13 $\sqrt{x^3-x^2-(1/x)}$

Exercise 4.5.14 $100/(100-x^2)^{3/2}$

Exercise 4.5.15 $\sqrt[3]{x+x^3}$

Exercise 4.5.16 $\sqrt{(x^2+1)^2+\sqrt{1+(x^2+1)^2}}$

Exercise 4.5.17 $(x+8)^5$

Exercise 4.5.18 $(4-x)^3$

Exercise 4.5.19 $(x^2+5)^3$

Exercise 4.5.20 $(6-2x^2)^3$

Exercise 4.5.21 $(1-4x^3)^{-2}$

Exercise 4.5.22 $5(x+1-1/x)$

Exercise 4.5.23 $4(2x^2-x+3)^{-2}$

Exercise 4.5.24 $\dfrac{1}{1+1/x}$

Exercise 4.5.25 $\dfrac{-3}{4x^2-2x+1}$

Exercise 4.5.26 $(x^2+1)(5-2x)/2$

Exercise 4.5.27 $(3x^2+1)(2x-4)^3$

Exercise 4.5.28 $\dfrac{x+1}{x-1}$

Exercise 4.5.29 $\dfrac{x^2-1}{x^2+1}$

Exercise 4.5.30 $\dfrac{(x-1)(x-2)}{x-3}$

Exercise 4.5.31 $\dfrac{2x^{-1}-x^{-2}}{3x^{-1}-4x^{-2}}$

Exercise 4.5.32 $3(x^2+1)(2x^2-1)(2x+3)$

Exercise 4.5.33 $\dfrac{1}{(2x+1)(x-3)}$

Exercise 4.5.34 $((2x+1)^{-1}+3)^{-1}$

Exercise 4.5.35 $(2x+1)^3(x^2+1)^2$

Exercise 4.5.36 *Find an equation for the tangent line to* $f(x)=(x-2)^{1/3}/(x^3+4x-1)^2$ *at* $x=1$.

Exercise 4.5.37 *Find an equation for the tangent line to* $y=9x^{-2}$ *at* $(3,1)$.

Exercise 4.5.38 *Find an equation for the tangent line to* $(x^2-4x+5)\sqrt{25-x^2}$ *at* $(3,8)$.

Exercise 4.5.39 *Find an equation for the tangent line to* $\dfrac{(x^2+x+1)}{(1-x)}$ *at* $(2,-7)$.

Exercise 4.5.40 *Find an equation for the tangent line to* $\sqrt{(x^2+1)^2+\sqrt{1+(x^2+1)^2}}$ *at* $(1,\sqrt{4+\sqrt{5}})$.

Exercise 4.5.41 *Let* $y=f(x)$ *and* $x=g(t)$. *If* $g(1)=2$, $f(2)=3$, $g'(1)=4$ *and* $f'(2)=5$, *find the derivative of* $f\circ g$ *at 1.*

Exercise 4.5.42 *Express the derivative of* $g(x)=x^2f(x^2)$ *in terms of* f *and the derivative of* f.

4.6 Derivatives of Exponential & Logarithmic Functions

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

$$\frac{d}{dx}a^x = \lim_{\Delta x\to 0}\frac{a^{x+\Delta x}-a^x}{\Delta x}$$

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

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

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

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

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

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

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

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

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

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

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

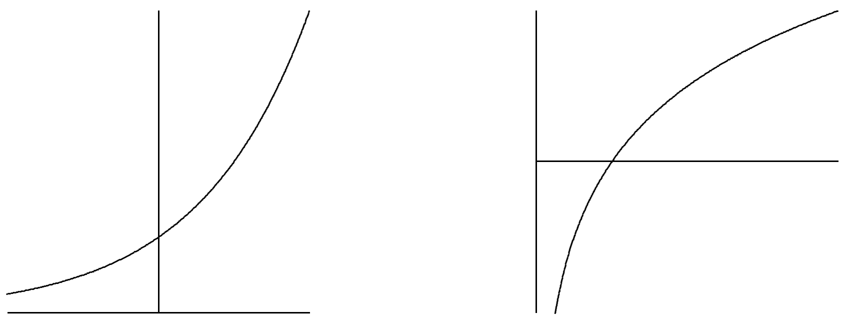


Figure 4.3: The exponential and logarithmic functions.

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

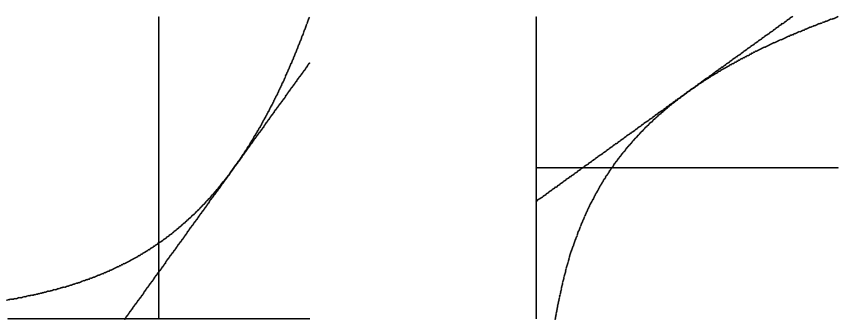


Figure 4.4: The exponential and logarithmic functions.

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

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

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

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

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

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

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

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

and we can compute the derivative using the chain rule:

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

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

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

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

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

Then

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

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

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

we can replace $\log_a e$ to get

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

You may if you wish memorize the formulas.

Derivative Formulas for a^x and $\log_a x$

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

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

Example 4.35: Derivative of Exponential Function

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

Solution.

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

♣

Example 4.36: Derivative of Exponential Function

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

Solution.

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

♣

Example 4.37: Power Rule

Recall that we have not justified the power rule except when the exponent is a positive or negative integer.

Solution. We can use the exponential function to take care of other exponents.

$$\frac{d}{dx}x^r = \frac{d}{dx}e^{r\ln x}$$

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

♣

Exercises for Section 4.6

Find the derivatives of the functions.

Exercise 4.6.1 3^{x^2}

Exercise 4.6.2 $\frac{\sin x}{e^x}$

Exercise 4.6.3 $(e^x)^2$

Exercise 4.6.4 $\sin(e^x)$

Exercise 4.6.5 $e^{\sin x}$

Exercise 4.6.6 $x^{\sin x}$

Exercise 4.6.7 $x^3 e^x$

Exercise 4.6.8 $x + 2^x$

Exercise 4.6.9 $(1/3)^{x^2}$

Exercise 4.6.10 e^{4x}/x

Exercise 4.6.11 $\ln(x^3 + 3x)$

Exercise 4.6.12 $\ln(\cos(x))$

Exercise 4.6.13 $\sqrt{\ln(x^2)}/x$

Exercise 4.6.14 $\ln(\sec(x) + \tan(x))$

Exercise 4.6.15 $x^{\cos(x)}$

Exercise 4.6.16 $x\ln x$

Exercise 4.6.17 $\ln(\ln(3x))$

Exercise 4.6.18 $\dfrac{1+\ln(3x^2)}{1+\ln(4x)}$

Exercise 4.6.19 *Find the value of a so that the tangent line to $y=\ln(x)$ at $x=a$ is a line through the origin. Sketch the resulting situation.*

Exercise 4.6.20 *If $f(x)=\ln(x^3+2)$ compute $f'(e^{1/3})$.*

4.7 Implicit Differentiation

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

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

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

Then using the chain rule on the right hand side:

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

Then we can solve for y':

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

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

The example $y=\ln x$ involved an inverse function defined implicitly, but other functions can be defined implicitly, and sometimes a single equation can be used to implicitly define more than one function.

Here's a familiar example.

Example 4.38: Derivative of Circle Equation

The equation $r^2 = x^2 + y^2$ describes a circle of radius r. The circle is not a function $y = f(x)$ because for some values of x there are two corresponding values of y. If we want to work with a function, we can break the circle into two pieces, the upper and lower semicircles, each of which is a function. Let's call these $y = U(x)$ and $y = L(x)$; in fact this is a fairly simple example, and it's possible to give explicit expressions for these: $U(x) = \sqrt{r^2 - x^2}$ and $L(x) = -\sqrt{r^2 - x^2}$. But it's somewhat easier, and quite useful, to view both functions as given implicitly by $r^2 = x^2 + y^2$: both $r^2 = x^2 + U(x)^2$ and $r^2 = x^2 + L(x)^2$ are true, and we can think of $r^2 = x^2 + y^2$ as defining both $U(x)$ and $L(x)$.
Now we can take the derivative of both sides as before, remembering that y is not simply a variable but a function—in this case, y is either $U(x)$ or $L(x)$ but we're not yet specifying which one. When we take the derivative we just have to remember to apply the chain rule where y appears.

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

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

Example 4.39: Slope of the Circle

Find the slope of the circle $4 = x^2 + y^2$ at the point $(1, -\sqrt{3})$.

Solution. Since we know both the x and y coordinates of the point of interest, we do not need to explicitly recognize that this point is on $L(x)$, and we do not need to use $L(x)$ to compute y – but we could. Using the calculation of y' from above,

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

It is instructive to compare this approach to others.

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

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

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

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

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

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

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

Example 4.40: Derivative of Function defined Implicitly

Find the derivative of any function defined implicitly by $yx^2+y^2=x$.

Solution. We treat y as an unspecified function and use the chain rule:

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

♣

Example 4.41: Derivative of Function defined Implicitly

Find the derivative of any function defined implicitly by $yx^2+e^y=x$.

Solution. We treat y as an unspecified function and use the chain rule:

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

♣

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

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

Example 4.42: Equation and Derivative of Ellipse

Discuss the equation and derivative of the ellispe.

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

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

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

Example 4.43: Derivative of Function defined Implicitly

Find $\frac{dy}{dx}$ by implicit differentiation if

$$2x^3+x^2y-y^9=3x+4.$$

Solution. Differentiating both sides with respect to x gives:

$$6x^2+\left(2xy+x^2\frac{dy}{dx}\right)-9y^8\frac{dy}{dx}=3,$$

$$x^2\frac{dy}{dx}-9y^8\frac{dy}{dx}=3-6x^2-2xy$$

$$\left(x^2-9y^8\right)\frac{dy}{dx}=3-6x^2-2xy$$

$$\frac{dy}{dx}=\frac{3-6x^2-2xy}{x^2-9y^8}.$$

♣

In the previous examples we had functions involving x and y, and we thought of y as a function of x. In these problems we differentiated with respect to x. So when faced with x's in the function we differentiated as usual, but when faced with y's we differentiated as usual except we multiplied by a $\frac{dy}{dx}$ for that term because we were using Chain Rule.

In the following example we will assume that both x and y are functions of t and want to differentiate the equation with respect to t. This means that every time we differentiate an x we will be using the Chain Rule, so we must multiply by $\frac{dx}{dt}$, and whenever we differentiate a y we multiply by $\frac{dy}{dt}$.

Example 4.44: Derivative of Function of an Additional Variable

Thinking of x and y as functions of t, differentiate the following equation with respect to t:

$$x^2 + y^2 = 100.$$

Solution. Using the Chain Rule we have:

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0.$$

♣

Example 4.45: Derivative of Function of an Additional Variable

If $y = x^3 + 5x$ and $\dfrac{dx}{dt} = 7$, find $\dfrac{dy}{dt}$ when $x = 1$.

Solution. Differentiating each side of the equation $y = x^3 + 5x$ with respect to t gives:

$$\frac{dy}{dt} = 3x^2\frac{dx}{dt} + 5\frac{dx}{dt}.$$

When $x = 1$ and $\frac{dx}{dt} = 7$ we have:

$$\frac{dy}{dt} = 3(1^2)(7) + 5(7) = 21 + 35 = 56.$$

♣

Logarithmic Differentiation

Previously we've seen how to do the derivative of a number to a function $(a^{f(x)})'$, and also a function to a number $[(f(x))^n]'$. But what about the derivative of a function to a function $[(f(x))^{g(x)}]'$?

In this case, we use a procedure known as **logarithmic differentiation**.

Steps for Logarithmic Differentiation

- Take ln of both sides of $y = f(x)$ to get $\ln y = \ln f(x)$ and simplify using logarithm properties,
- Differentiate implicitly with respect to x and solve for $\dfrac{dy}{dx}$,
- Replace y with its function of x (i.e., $f(x)$).

Example 4.46: Logarithmic Differentiation

Differentiate $y = x^x$.

Solution. We take ln of both sides:

$$\ln y = \ln x^x.$$

Using log properties we have:

$$\ln y = x \ln x.$$

Differentiating implicitly gives:

$$\frac{y'}{y} = (1)\ln x + x\frac{1}{x}.$$

$$\frac{y'}{y} = \ln x + 1.$$

Solving for y' gives:

$$y' = y(1 + \ln x).$$

Replace $y = x^x$ gives:

$$y' = x^x(1 + \ln x).$$

Another method to find this derivative is as follows:

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

♣

In fact, logarithmic differentiation can be used on more complicated products and quotients (not just when dealing with functions to the power of functions).

Example 4.47: Logarithmic Differentiation

Differentiate (assuming $x > 0$):

$$y = \frac{(x+2)^3(2x+1)^9}{x^8(3x+1)^4}.$$

Solution. Using product & quotient rules for this problem is a complete nightmare! Let's apply logarithmic differentiation instead. Take ln of both sides:

$$\ln y = \ln\left(\frac{(x+2)^3(2x+1)^9}{x^8(3x+1)^4}\right).$$

Applying log properties:

$$\ln y = \ln\left((x+2)^3(2x+1)^9\right) - \ln\left(x^8(3x+1)^4\right).$$

$$\ln y = \ln\left((x+2)^3\right) + \ln\left((2x+1)^9\right) - \left[\ln\left(x^8\right) + \ln\left((3x+1)^4\right)\right].$$
$$\ln y = 3\ln(x+2) + 9\ln(2x+1) - 8\ln x - 4\ln(3x+1).$$

Now, differentiating implicitly with respect to x gives:

$$\frac{y'}{y} = \frac{3}{x+2} + \frac{18}{2x+1} - \frac{8}{x} - \frac{12}{3x+1}.$$

Solving for y' gives:

$$y' = y\left(\frac{3}{x+2} + \frac{18}{2x+1} - \frac{8}{x} - \frac{12}{3x+1}\right).$$

Replace $y = \frac{(x+2)^3(2x+1)^9}{x^8(3x+1)^4}$ gives:

$$y' = \frac{(x+2)^3(2x+1)^9}{x^8(3x+1)^4}\left(\frac{3}{x+2} + \frac{18}{2x+1} - \frac{8}{x} - \frac{12}{3x+1}\right).$$

♣

Exercises for Section 4.7

Exercise 4.7.1 *Find a formula for the derivative y' at the point (x,y):*

(a) $y^2 = 1 + x^2$

(b) $x^2 + xy + y^2 = 7$

(c) $x^3 + xy^2 = y^3 + yx^2$

(d) $4\cos x \sin y = 1$

(e) $\sqrt{x} + \sqrt{y} = 9$

(f) $\tan(x/y) = x + y$

(g) $\sin(x+y) = xy$

(h) $\frac{1}{x} + \frac{1}{y} = 7$

Exercise 4.7.2 *A hyperbola passing through $(8,6)$ consists of all points whose distance from the origin is a constant more than its distance from the point (5,2). Find the slope of the tangent line to the hyperbola at $(8,6)$.*

Exercise 4.7.3 *The graph of the equation $x^2 - xy + y^2 = 9$ is an ellipse. Find the lines tangent to this curve at the two points where it intersects the x-axis. Show that these lines are parallel.*

Exercise 4.7.4 *Repeat the previous problem for the points at which the ellipse intersects the y-axis.*

Exercise 4.7.5 *If $y = \log_a x$ then $a^y = x$. Use implicit differentiation to find y'.*

Exercise 4.7.6 *Find the points on the ellipse from the previous two problems where the slope is horizontal and where it is vertical.*

Exercise 4.7.7 *Find an equation for the tangent line to $x^4 = y^2 + x^2$ at $(2, \sqrt{12})$. (This curve is the* **kampyle of Eudoxus**.*)*

Exercise 4.7.8 *Find an equation for the tangent line to $x^{2/3} + y^{2/3} = a^{2/3}$ at a point (x_1, y_1) on the curve, with $x_1 \neq 0$ and $y_1 \neq 0$. (This curve is an* **a***stroid.)*

Exercise 4.7.9 *Find an equation for the tangent line to $(x^2+y^2)^2 = x^2 - y^2$ at a point (x_1, y_1) on the curve, with $x_1 \neq 0, -1, 1$. (This curve is a* **lemniscate**.*)*

Exercise 4.7.10 *Two curves are* **orthogonal** *if at each point of intersection, the angle between their tangent lines is $\pi/2$. Two families of curves, $\mathscr{A}$ and $\mathscr{B}$, are* **orthogonal trajectories** *of each other if given any curve C in $\mathscr{A}$ and any curve D in $\mathscr{B}$ the curves C and D are orthogonal. For example, the family of horizontal lines in the plane is orthogonal to the family of vertical lines in the plane.*

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

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

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

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

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

Exercise 4.7.11 *Differentiate the function $y = \dfrac{(x-1)^8(x-23)^{1/2}}{27x^6(4x-6)^8}$*

Exercise 4.7.12 *Differentiate the function $f(x) = (x+1)^{\sin x}$.*

Exercise 4.7.13 *Differentiate the function $g(x) = \dfrac{e^x(\cos x + 2)^3}{\sqrt{x^2+4}}$.*

4.8 Derivatives of Inverse Functions

Suppose we wanted to find the *derivative of the inverse*, but do not have an actual formula for the inverse function? Then we can use the following derivative formula for the inverse evaluated at a.

Derivative of $f^{-1}(a)$

Given an invertible function $f(x)$, the derivative of its inverse function $f^{-1}(x)$ evaluated at $x = a$ is:

$$\left[f^{-1}\right]'(a) = \frac{1}{f'\left[f^{-1}(a)\right]}$$

To see why this is true, start with the function $y = f^{-1}(x)$. Write this as $x = f(y)$ and differentiate both sides implicitly with respect to x using the chain rule:

$$1 = f'(y) \cdot \frac{dy}{dx}.$$

Thus,

$$\frac{dy}{dx} = \frac{1}{f'(y)},$$

but $y = f^{-1}(x)$, thus,

$$\left[f^{-1}\right]'(x) = \frac{1}{f'\left[f^{-1}(x)\right]}.$$

At the point $x = a$ this becomes:

$$\left[f^{-1}\right]'(a) = \frac{1}{f'\left[f^{-1}(a)\right]}$$

Example 4.48: Derivatives of Inverse Functions

Suppose $f(x) = x^5 + 2x^3 + 7x + 1$. *Find* $\left[f^{-1}\right]'(1)$.

Solution. First we should show that f^{-1} exists (i.e. that f is one-to-one). In this case the derivative $f'(x) = 5x^4 + 6x^2 + 7$ is strictly greater than 0 for all x, so f is strictly increasing and thus one-to-one.

It's difficult to find the inverse of $f(x)$ (and then take the derivative). Thus, we use the above formula evaluated at 1:

$$\left[f^{-1}\right]'(1) = \frac{1}{f'\left[f^{-1}(1)\right]}.$$

Note that to use this formula we need to know what $f^{-1}(1)$ is, and the derivative $f'(x)$. To find $f^{-1}(1)$ we make a table of values (plugging in $x = -3, -2, -1, 0, 1, 2, 3$ into $f(x)$) and see what value of x gives 1. We omit the table and simply observe that $f(0) = 1$. Thus,

$$f^{-1}(1) = 0.$$

Now we have:

$$\left[f^{-1}\right]'(1) = \frac{1}{f'(0)}.$$

And so, $f'(0) = 7$. Therefore,

$$\left[f^{-1}\right]'(1) = \frac{1}{7}.$$

♣

4.8.1. Derivatives of Inverse Trigonometric Functions

We can apply the technique used to find the derivative of f^{-1} above to find the derivatives of the inverse trigonometric functions.

In the following examples we will derive the formulae for the derivative of the inverse sine, inverse cosine and inverse tangent. The other three inverse trigonometric functions have been left as exercises at the end of this section.

Example 4.49: Derivative of Inverse Sine

Find the derivative of $\sin^{-1}(x)$.

Solution. As above, we write $y = \sin^{-1}(x)$, so $x = \sin(y)$ and $-\pi/2 \leq y \leq \pi/2$, and differentiate both sides with respect to x using the chain rule.

$$\begin{aligned}
\frac{d}{dx}x &= \frac{d}{dx}\sin(y) \\
1 &= \cos(y)\frac{dy}{dx} \\
1 &= \cos\left(\sin^{-1}(x)\right)\frac{dy}{dx} \\
\frac{dy}{dx} &= \frac{1}{\cos\left(\sin^{-1}(x)\right)}
\end{aligned}$$

Although correct, this formula is cumbersome to use, and can be simplified significantly with a bit of trigonometry. Let $\theta = \sin^{-1}(x)$, so $\sin(\theta) = x$, and construct a right angle triangle with angle θ, opposite side length x and hypotenuse 1. The Pythagorean Theorem gives an adjacent side length of $\sqrt{1-x^2}$, so $\cos\left(\sin^{-1}(x)\right) = \cos(\theta) = \sqrt{1-x^2}$. Note that we choose the non-negative square root $\sqrt{1-x^2}$ since $\cos(\theta) \geq 0$ when $-\pi/2 \leq \theta \leq \pi/2$.

Finally, the derivative of inverse sine is

$$\left(\sin^{-1}(x)\right)' = \frac{1}{\sqrt{1-x^2}}$$

♣

Example 4.50: Derivative of Inverse Cosine

Find the derivative of $\cos^{-1}(x)$.

Solution. Let $y = \cos^{-1}(x)$, so $\cos(y) = x$ and $0 \le y \le \pi$. Next we differentiate implicitly:

$$\begin{aligned} \frac{d}{dx}(\cos y) &= \frac{d}{dx}(x) \\ -\sin y \cdot \frac{dy}{dx} &= 1 \\ \frac{dy}{dx} &= -\frac{1}{\sin y} \end{aligned}$$

This time, just for variety, we will leave the derivative in terms of y and apply some trigonometry. Since $\cos y = x$, we construct a triangle with angle y, adjacent side length x and hypotenuse 1. Solving for the opposite side length using Pythagorean Theorem we obtain $\sqrt{1-x^2}$. Using this triangle we can see that $\sin y = \sqrt{1-x^2}$ $(0 \le y \le \pi)$. Substituting this into the equation for dy/dx, we find that

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

. ♣

In the following example we explore an alternate method of finding the derivative.

Example 4.51: Derivative of Inverse Tangent

Find the derivative of $\tan^{-1}(x)$.

Solution. We begin with $\tan\left(\tan^{-1}(x)\right) = x$. Taking the derivative using the Chain Rule we obtain

$$\sec^2\left(\tan^{-1}(x)\right) \cdot \frac{d}{dx}\left(\tan^{-1}(x)\right) = 1,$$

which we rearrange to obtain

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

Let $\tan^{-1}(x) = \theta$, then $\tan(\theta) = x$. We construct a triangle with angle θ, adjacent side 1 and opposite side x. The hypotenuse is $\sqrt{1+x^2}$ using Pythagorean theorem. Then $\sec^2\left(\tan^{-1}(x)\right) = \sec^2(\theta) = (\sec(\theta))^2 = \left(\sqrt{1+x^2}\right)^2 = 1+x^2$. Recall that $\sec(x) = 1/\cos(x)$. Finally, the derivative is

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

♣

Exercises for 4.8

Exercise 4.8.1 *Given $f(x) = 1 + \ln(x-2)$, first show that f^{-1} exists, then compute $\left[f^{-1}\right]'(1)$.*

Exercise 4.8.2 *The* **inverse cotangent function**, *denoted by* $\cot^{-1}(x)$, *is defined to be the inverse of the restricted cotangent function:* $\cot(x)$, $0 < x < \pi$. *Find the derivative of* $\cot^{-1}(x)$.

Exercise 4.8.3 *The* **inverse secant function**, *denoted by* $\sec^{-1}(x)$, *is defined to be the inverse of the restricted secant function:* $\sec(x)$, $x \in [0, \pi/2) \cup [\pi, 3\pi/2)$. *Find the derivative of* $\sec^{-1}(x)$.

Exercise 4.8.4 *The* **inverse cosecant function**, *denoted by* $\csc^{-1}(x)$, *is defined to be the inverse of the restricted cosecant function:* $\csc(x)$, $x \in (0, \pi/2] \cup (\pi, 3\pi/2]$. *Find the derivative of* $\csc^{-1}(x)$.

Exercise 4.8.5 *Suppose $f(x) = x^3 + 4x + 2$. Find the slope of the tangent line to the graph of $g(x) = xf^{-1}(x)$ at the point where $x = 7$.*

Exercise 4.8.6 *Find the derivatives of* $\sin^{-1}(x) + \cos^{-1}(x)$ *and* $(x^2+1)\tan^{-1}(x)$.

Exercise 4.8.7 *Differentiate $y = \sin^{-1}(x^2)$ and $y = \tan^{-1}(3x)$.*

4.9 Additional Exercises

Exercise 4.9.1 *Find the derivatives of the following functions from definition.*

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

(b) $g(x) = x^{3/2}$

Exercise 4.9.2 *Let* $f(x) = \begin{cases} x^3 & \text{if } x \leq 1 \\ 5x - x^2 & \text{if } x > 1 \end{cases}$. *Use the definition of the derivative to find $f'(1)$.*

Exercise 4.9.3 *Differentiate the following functions.*

(a) $y = 7x^4 - 7\pi^4 + \dfrac{1}{\pi\sqrt[3]{x}}$

(b) $f(x) = \dfrac{1-\sqrt{x}}{1+\sqrt{x}}$

(c) $f(x) = |x-1| + |x+2|$

(d) $f(x) = x^2 \sin x \cos x$

(e) $y = \dfrac{x \sin x}{1 + \sin x}$

(f) $g(x) = \sqrt{2 + \dfrac{3}{\sqrt{x}}}$

(g) $y = \sqrt[3]{x^4 + x^2 + 1} + \dfrac{1}{(x^3 - x + 4)^5}$

(h) $y = \sin^3 x - \sin(x^3)$

(i) $F(x) = \sec^4 x + \tan^4 x$

(j) $y = \cos^2 \left(\dfrac{1-x}{1+x} \right)$

(k) $y = \tan(\sin(x^2 + \sec^2 x))$

(l) $y = \dfrac{1}{2 + \sin \frac{\pi}{x}}$

Exercise 4.9.4 *Differentiate the following functions.*

(a) $y = e^{3x} + e^{-x} + e^2$

(b) $y = e^{2x} \cos 3x$

(c) $f(x) = \tan(x + e^x)$

(d) $g(x) = \dfrac{e^x}{e^x + 2}$

(e) $y = \ln(2 + \sin x) - \sin(2 + \ln x)$

(f) $f(x) = e^{x^\pi} + x^{\pi^e} + \pi^{e^x}$

(g) $y = \log_a(b^x) + b^{\log_a x}$, *where a and b are positive real numbers and* $a \neq 1$

(h) $y = (x^2 + 1)^{x^3+1}$

(i) $y = (x^2 + e^x)^{1/\ln x}$

(j) $y = \dfrac{x\sqrt{x^2 + x + 1}}{(2 + \sin x)^4 (3x + 5)^7}$

Exercise 4.9.5 *Find* $\dfrac{dy}{dx}$ *if y is a differentiable function that satisfy the given equation.*

(a) $x^2 + xy + y^2 = 7$

(b) $x^2 + y^2 = (2x^2 + 2y^2 - x)^2$

(c) $x^2 \sin y + y^3 = \cos x$

(d) $x^2 + xe^y = 2y + e^x$

Exercise 4.9.6 *Differentiate the following functions.*

(a) $y = x \sin^{-1} x$

(b) $f(x) = \dfrac{\sin^{-1} x}{\cos^{-1} x}$

(c) $g(x) = \tan^{-1}\left(\dfrac{x}{a}\right)$, *where* $a > 0$

(d) $y = x \tan^{-1} x - \dfrac{1}{2}\ln(x^2 + 1)$

5. Applications of Derivatives

In this chapter we explore how to use derivative and differentiation to solve a variety of problems, some mathematical and some practical. We explore some applications which motivated and were formalized in the definition of the derivative, and look at a few clever uses of the tangent line (which has immediate geometric ties to the definition of the derivative).

5.1 Related Rates

When defining the derivative $f'(x)$, we define it to be exactly the rate of change of $f(x)$ with respect to x. Consequently, any question about rates of change can be rephrased as a question about derivatives. **When we calculate derivatives, we are calculating rates of change**. Results and answers we obtain for derivatives translate directly into results and answers about rates of change. Let us look at some examples where more than one variable is involved, and where our job is to analyze and exploit relations between the rates of change of these variables. The mathematical step of relating the rates of change turns out to be largely an exercise in differentiation using the chain rule or implicit differentiation. This explains why some textbooks place this section shortly after the sections on the chain rule and implicit differentiation.

Suppose we have two variables x and y (in most problems the letters will be different, but for now let's use x and y) which are both changing with time. A "related rates" problem is a problem in which we know one of the rates of change at a given instant—say, $\dot{x} = dx/dt$—and we want to find the other rate $\dot{y} = dy/dt$ at that instant. (The use of $\dot{x}$ to mean dx/dt goes back to Newton and is still used for this purpose, especially by physicists.)

If y is written in terms of x, i.e., $y = f(x)$, then this is easy to do using the chain rule:

$$\dot{y} = \frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt} = \frac{dy}{dx}\dot{x}.$$

That is, find the derivative of $f(x)$, plug in the value of x at the instant in question, and multiply by the given value of $\dot{x} = dx/dt$ to get $\dot{y} = dy/dt$.

Example 5.1: Speed at which a Coordinate is Changing

Suppose an object is moving along a path described by $y = x^2$, that is, it is moving on a parabolic path. At a particular time, say $t = 5$, the x coordinate is 6 and we measure the speed at which the x coordinate of the object is changing and find that $dx/dt = 3$.
At the same time, how fast is the y coordinate changing?

Solution. Using the chain rule, $dy/dt = 2x \cdot dx/dt$. At $t = 5$ we know that $x = 6$ and $dx/dt = 3$, so $dy/dt = 2 \cdot 6 \cdot 3 = 36$. ♣

In many cases, particularly interesting ones, x and y will be related in some other way, for example $x = f(y)$, or $F(x,y) = k$, or perhaps $F(x,y) = G(x,y)$, where $F(x,y)$ and $G(x,y)$ are expressions involving both variables. In all cases, you can solve the related rates problem by taking the derivative of both sides, plugging in all the known values (namely, x, y, and $\dot{x}$), and then solving for $\dot{y}$.

To summarize, here are the steps in doing a related rates problem.

Steps for Solving Related Rates Problems

1. Decide what the two variables are.
2. Find an equation relating them.
3. Take d/dt of both sides.
4. Plug in all known values at the instant in question.
5. Solve for the unknown rate.

Example 5.2: Receding Airplanes

A plane is flying directly away from you at 500 mph at an altitude of 3 miles. How fast is the plane's distance from you increasing at the moment when the plane is flying over a point on the ground 4 miles from you?

Solution. To see what's going on, we first draw a schematic representation of the situation, as in Figure 5.1.

Because the plane is in level flight directly away from you, the rate at which x changes is the speed of the plane, $dx/dt = 500$. The distance between you and the plane is y; it is dy/dt that we wish to know. By the Pythagorean Theorem we know that $x^2 + 9 = y^2$. Taking the derivative:

$$2x\dot{x} = 2y\dot{y}.$$

We are interested in the time at which $x = 4$; at this time we know that $4^2 + 9 = y^2$, so $y = 5$. Putting together all the information we get

$$2(4)(500) = 2(5)\dot{y}.$$

Thus, $\dot{y} = 400$ mph. ♣

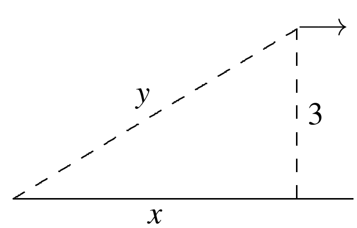


Figure 5.1: Receding airplane.

Example 5.3: Spherical Balloon

You are inflating a spherical balloon at the rate of 7 cm^3/sec. How fast is its radius increasing when the radius is 4 cm?

Solution. Here the variables are the radius r and the volume V. We know dV/dt, and we want dr/dt. The two variables are related by the equation $V = 4\pi r^3/3$. Taking the derivative of both sides gives $dV/dt = 4\pi r^2 \dot{r}$. We now substitute the values we know at the instant in question: $7 = 4\pi 4^2 \dot{r}$, so $\dot{r} = 7/(64\pi)$ cm/sec. ♣

Example 5.4: Conical Container

Water is poured into a conical container at the rate of 10 cm^3/sec. The cone points directly down, and it has a height of 30 cm and a base radius of 10 cm; see Figure 5.2. How fast is the water level rising when the water is 4 cm deep (at its deepest point)?

Solution. The water forms a conical shape within the big cone; its height and base radius and volume are all increasing as water is poured into the container. This means that we actually have three things varying with time: the water level h (the height of the cone of water), the radius r of the circular top surface of water (the base radius of the cone of water), and the volume of water V. The volume of a cone is given by $V = \pi r^2 h/3$. We know dV/dt, and we want dh/dt. At first something seems to be wrong: we have a third variable, r, whose rate we don't know.

However, the dimensions of the cone of water must have the same proportions as those of the container. That is, because of similar triangles, $r/h = 10/30$ so $r = h/3$. Now we can eliminate r from the problem entirely: $V = \pi(h/3)^2 h/3 = \pi h^3/27$. We take the derivative of both sides and plug in $h = 4$ and $dV/dt = 10$, obtaining $10 = (3\pi \cdot 4^2/27)(dh/dt)$. Thus, $dh/dt = 90/(16\pi)$ cm/sec. ♣

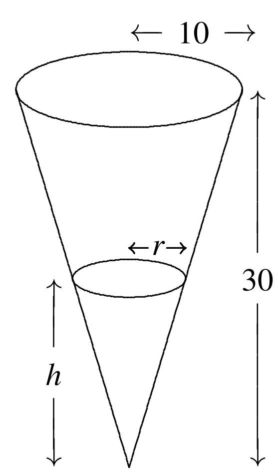


Figure 5.2: Conical water tank.

Example 5.5: Swing Set

A swing consists of a board at the end of a 10 ft long rope. Think of the board as a point P at the end of the rope, and let Q be the point of attachment at the other end. Suppose that the swing is directly below Q at time $t = 0$, and is being pushed by someone who walks at 6 ft/sec from left to right. Find (a) how fast the swing is rising after 1 sec; (b) the angular speed of the rope in deg/sec after 1 sec.

Solution. We start out by asking: What is the geometric quantity whose rate of change we know, and what is the geometric quantity whose rate of change we're being asked about? Note that the person pushing the swing is moving horizontally at a rate we know. In other words, the horizontal coordinate of P is increasing at 6 ft/sec. In the xy-plane let us make the convenient choice of putting the origin at the location of P at time $t = 0$, i.e., a distance 10 directly below the point of attachment. Then the rate we know is dx/dt, and in part (a) the rate we want is dy/dt (the rate at which P is rising). In part (b) the rate we want is $\dot{\theta} = d\theta/dt$, where θ stands for the angle in radians through which the swing has swung from the vertical. (Actually, since we want our answer in deg/sec, at the end we must convert $d\theta/dt$ from rad/sec by multiplying by $180/\pi$.)

(a) From the diagram we see that we have a right triangle whose legs are x and $10-y$, and whose hypotenuse is 10. Hence $x^2 + (10-y)^2 = 100$. Taking the derivative of both sides we obtain: $2x\dot{x} + 2(10-y)(0-\dot{y}) = 0$. We now look at what we know after 1 second, namely $x = 6$ (because x started at 0 and has been increasing at the rate of 6 ft/sec for 1 sec), thus $y = 2$ (because we get $10 - y = 8$ from the Pythagorean theorem applied to the triangle with hypotenuse 10 and leg 6), and $\dot{x} = 6$. Putting in these values gives us $2 \cdot 6 \cdot 6 - 2 \cdot 8\dot{y} = 0$, from which we can easily solve for $\dot{y}$: $\dot{y} = 4.5$ ft/sec.

(b) Here our two variables are x and θ, so we want to use the same right triangle as in part (a), but this time relate θ to x. Since the hypotenuse is constant (equal to 10), the best way to do this is to use the sine: $\sin\theta = x/10$. Taking derivatives we obtain $(\cos\theta)\dot{\theta} = 0.1\dot{x}$. At the instant in question ($t = 1$ sec), when we have a right triangle with sides 6–8–10, $\cos\theta = 8/10$ and $\dot{x} = 6$. Thus $(8/10)\dot{\theta} = 6/10$, i.e., $\dot{\theta} = 6/8 = 3/4$ rad/sec, or approximately 43 deg/sec. ♣

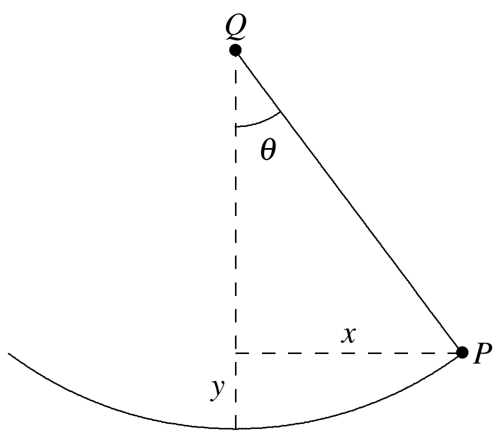


Figure 5.3: Swing.

We have seen that sometimes there are apparently more than two variables that change with time, but in reality there are just two, as the others can be expressed in terms of just two. However sometimes there really are several variables that change with time; as long as you know the rates of change of all but one of them you can find the rate of change of the remaining one. As in the case when there are just two variables, take the derivative of both sides of the equation relating all of the variables, and then substitute all of the known values and solve for the unknown rate.

Example 5.6: Distance Changing Rate

A road running north to south crosses a road going east to west at the point P. Car A is driving north along the first road, and car B is driving east along the second road. At a particular time car A is 10 kilometers to the north of P and traveling at 80 km/hr, while car B is 15 kilometers to the east of P and traveling at 100 km/hr. How fast is the distance between the two cars changing?

Solution. Let $a(t)$ be the distance of car A north of P at time t, and $b(t)$ the distance of car B east of P at time t, and let $c(t)$ be the distance from car A to car B at time t. By the Pythagorean Theorem, $c(t)^2 = a(t)^2 + b(t)^2$. Taking derivatives we get $2c(t)c'(t) = 2a(t)a'(t) + 2b(t)b'(t)$, so

$$\dot{c} = \frac{a\dot{a} + b\dot{b}}{c} = \frac{a\dot{a} + b\dot{b}}{\sqrt{a^2 + b^2}}.$$

Substituting known values we get:

$$\dot{c} = \frac{10 \cdot 80 + 15 \cdot 100}{\sqrt{10^2 + 15^2}} = \frac{460}{\sqrt{13}} \approx 127.6\text{km/hr}$$

at the time of interest. ♣

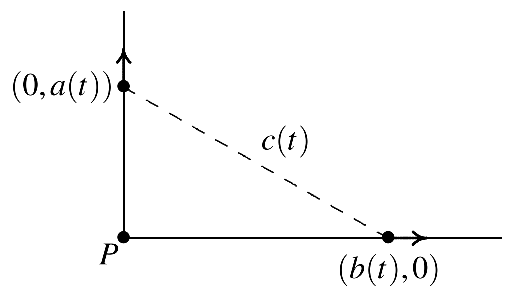


Figure 5.4: Cars moving apart.

Notice how this problem differs from Example 5.2. In both cases we started with the Pythagorean Theorem and took derivatives on both sides. However, in Example 5.2 one of the sides was a constant (the altitude of the plane), and so the derivative of the square of that side of the triangle was simply zero. In this Example, on the other hand, all three sides of the right triangle are variables, even though we are interested in a specific value of each side of the triangle (namely, when the sides have lengths 10 and 15). Make sure that you understand at the start of the problem what are the variables and what are the constants.

Exercises for Section 5.1

Exercise 5.1.1 *Air is being pumped into a spherical balloon at a constant rate of 3 cm^3/s. How fast is the radius of the balloon increasing when the radius reaches 5cm?*

Exercise 5.1.2 *A cylindrical tank standing upright (with one circular base on the ground) has radius 20 cm. How fast does the water level in the tank drop when the water is being drained at 25 cm^3/sec?*

Exercise 5.1.3 *A cylindrical tank standing upright (with one circular base on the ground) has radius 1 meter. How fast does the water level in the tank drop when the water is being drained at 3 liters per second?*

Exercise 5.1.4 *A ladder 13 meters long rests on horizontal ground and leans against a vertical wall. The foot of the ladder is pulled away from the wall at the rate of 0.6 m/sec. How fast is the top sliding down the wall when the foot of the ladder is 5 m from the wall?*

Exercise 5.1.5 *A ladder 13 meters long rests on horizontal ground and leans against a vertical wall. The top of the ladder is being pulled up the wall at* 0.1 *meters per second. How fast is the foot of the ladder approaching the wall when the foot of the ladder is 5 m from the wall?*

Exercise 5.1.6 *A rotating beacon is located 2 miles out in the water. Let A be the point on the shore that is closest to the beacon. As the beacon rotates at 10 rev/min, the beam of light sweeps down the shore once each time it revolves. Assume that the shore is straight. How fast is the point where the beam hits the shore moving at an instant when the beam is lighting up a point 2 miles along the shore from the point A?*

Exercise 5.1.7 *A baseball diamond is a square 90 ft on a side. A player runs from first base to second base at 15 ft/sec. At what rate is the player's distance from third base decreasing when she is half way from first to second base?*

Exercise 5.1.8 *Sand is poured onto a surface at 15 cm^3/sec, forming a conical pile whose base diameter is always equal to its altitude. How fast is the altitude of the pile increasing when the pile is 3 cm high?*

Exercise 5.1.9 *A boat is pulled in to a dock by a rope with one end attached to the front of the boat and the other end passing through a ring attached to the dock at a point 5 ft higher than the front of the boat. The rope is being pulled through the ring at the rate of 0.6 ft/sec. How fast is the boat approaching the dock when 13 ft of rope are out?*

Exercise 5.1.10 *A balloon is at a height of 50 meters, and is rising at the constant rate of 5 m/sec. A bicyclist passes beneath it, traveling in a straight line at the constant speed of 10 m/sec. How fast is the distance between the bicyclist and the balloon increasing 2 seconds later?*

Exercise 5.1.11 *A pyramid-shaped vat has square cross-section and stands on its tip. The dimensions at the top are 2 m $\times$ 2 m, and the depth is 5 m. If water is flowing into the vat at 3 m^3/min, how fast is the water level rising when the depth of water (at the deepest point) is 4 m? Note: the volume of any "conical" shape (including pyramids) is* $(1/3)(\textit{height})(\textit{area of base})$.

Exercise 5.1.12 *A woman 5 ft tall walks at the rate of 3.5 ft/sec away from a streetlight that is 12 ft above the ground. At what rate is the tip of her shadow moving? At what rate is her shadow lengthening?*

Exercise 5.1.13 *A man 1.8 meters tall walks at the rate of 1 meter per second toward a streetlight that is 4 meters above the ground. At what rate is the tip of his shadow moving? At what rate is his shadow shortening?*

Exercise 5.1.14 *A police helicopter is flying at 150 mph at a constant altitude of 0.5 mile above a straight road. The pilot uses radar to determine that an oncoming car is at a distance of exactly 1 mile from the helicopter, and that this distance is decreasing at 190 mph. Find the speed of the car.*

Exercise 5.1.15 *A police helicopter is flying at 200 kilometers per hour at a constant altitude of 1 km above a straight road. The pilot uses radar to determine that an oncoming car is at a distance of exactly 2 kilometers from the helicopter, and that this distance is decreasing at 250 kph. Find the speed of the car.*

Exercise 5.1.16 *A light shines from the top of a pole 20 m high. An object is dropped from the same height from a point 10 m away, so that its height at time t seconds is $h(t) = 20 - 9.8t^2/2$. How fast is the object's shadow moving on the ground one second later?*

5.2 Extrema of a Function

In calculus, there is much emphasis placed on analyzing the behaviour of a function f on an interval I. Does f have a maximum value on I? Does it have a minimum value? How does the interval I impact our discussion of extrema?

5.2.1. Local Extrema

A **local maximum** point on a function is a point (x,y) on the graph of the function whose y coordinate is larger than all other y coordinates on the graph at points "close to" (x,y). More precisely, $(x, f(x))$ is a local maximum if there is an interval (a,b) with $a < x < b$ and $f(x) \geq f(z)$ for every z in (a,b). Similarly, (x,y) is a **local minimum** point if it has locally the smallest y coordinate. Again being more precise: $(x, f(x))$ is a local minimum if there is an interval (a,b) with $a < x < b$ and $f(x) \leq f(z)$ for every z in (a,b). A **local extremum** is either a local minimum or a local maximum.

Definition 5.7: Local Maxima and Minima

A real-valued function f has a **local maximum** *at x_0 if $f(x_0)$ is the largest value of f near x_0; in other words, $f(x_0 \geq f(x)$ when x is near x_0.*

A real-valued function f has a **local minimum** *at x_0 if $f(x_0)$ is the smallest value of f near x_0; in other words, $f(x_0) \leq f(x)$ when x is near x_0.*

Local maximum and minimum points are quite distinctive on the graph of a function, and are therefore useful in understanding the shape of the graph. In many applied problems we want to find the largest or smallest value that a function achieves (for example, we might want to find the minimum cost at which some task can be performed) and so identifying maximum and minimum points will be useful for applied problems as well. Some examples of local maximum and minimum points are shown in Figure 5.5.

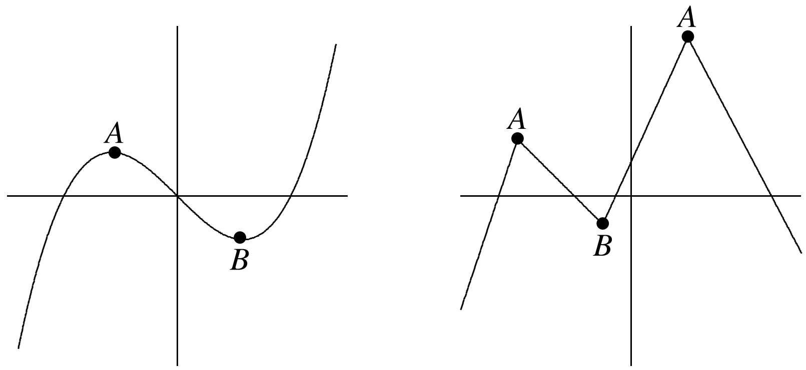


Figure 5.5: Some local maximum points (A) and minimum points (B).

If $(x, f(x))$ is a point where $f(x)$ reaches a local maximum or minimum, and if the derivative of f exists at x, then the graph has a tangent line and the tangent line *must* be horizontal. This is important enough to state as a theorem.

The proof is simple enough and we include it here, but you may accept Fermat's Theorem based on its strong intuitive appeal and come back to its proof at a later time.

Theorem 5.8: Fermat's Theorem

If $f(x)$ has a local extremum at $x = a$ and f is differentiable at a, then $f'(a) = 0$.

Proof. We shall give the proof for the case where $f(x)$ has a local maximum at $x = a$. The proof for the local minimum case is similar.

Since $f(x)$ has a local maximum at $x = a$, there is an open interval (c, d) with $c < a < d$ and $f(x) \leq f(a)$ for every x in (c, d). So, $f(x) - f(a) \leq 0$ for all such x. Let us now look at the sign of the difference quotient $\dfrac{f(x) - f(a)}{x - a}$. We consider two cases according as $x > a$ or $x < a$.

If $x > a$, then $x - a > 0$ and so, $\dfrac{f(x) - f(a)}{x - a} \leq 0$. Taking limit as x approach a from the right, we get

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} \leq 0.$$

On the other hand, if $x < a$, then $x - a < 0$ and so, $\dfrac{f(x) - f(a)}{x - a} \geq 0$. Taking limit as x approach a from the left, we get

$$\lim_{x \to a^-} \frac{f(x) - f(a)}{x - a} \geq 0.$$

Since f is differentiable at a, $f'(a) = \displaystyle\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a^-} \frac{f(x) - f(a)}{x - a}$. Therefore, we have both $f'(a) \leq 0$ and $f'(a) \geq 0$. So, $f'(a) = 0$. ♣

Thus, the only points at which a function can have a local maximum or minimum are points at which the derivative is zero, as in the left hand graph in Figure 5.5, or the derivative is undefined, as in the right hand graph. Any value of x in the domain of f for which $f'(x)$ is zero or undefined is called a **critical**

point for f. When looking for local maximum and minimum points, you are likely to make two sorts of mistakes: You may forget that a maximum or minimum can occur where the derivative does not exist, and so forget to check whether the derivative exists everywhere. You might also assume that any place that the derivative is zero is a local maximum or minimum point, but this is not true. A portion of the graph of $f(x) = x^3$ is shown in Figure 5.6. The derivative of f is $f'(x) = 3x^2$, and $f'(0) = 0$, but there is neither a maximum nor minimum at $(0,0)$.

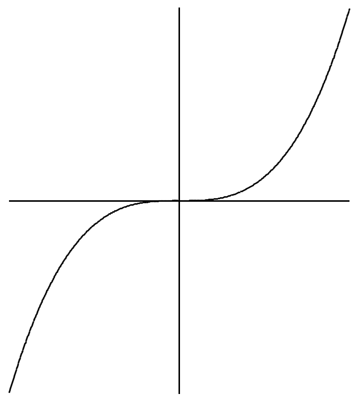


Figure 5.6: No maximum or minimum even though the derivative is zero.

Since the derivative is zero or undefined at both local maximum and local minimum points, we need a way to determine which, if either, actually occurs. The most elementary approach, but one that is often tedious or difficult, is to test directly whether the y coordinates "near" the potential maximum or minimum are above or below the y coordinate at the point of interest. Of course, there are too many points "near" the point to test, but a little thought shows we need only test two provided we know that f is continuous (recall that this means that the graph of f has no jumps or gaps).

Suppose, for example, that we have identified three points at which f' is zero or nonexistent: (x_1,y_1), (x_2,y_2), (x_3,y_3), and $x_1 < x_2 < x_3$ (see Figure 5.7). Suppose that we compute the value of $f(a)$ for $x_1 < a < x_2$, and that $f(a) < f(x_2)$. What can we say about the graph between a and x_2? Could there be a point $(b, f(b))$, $a < b < x_2$ with $f(b) > f(x_2)$? No: if there were, the graph would go up from $(a, f(a))$ to $(b, f(b))$ then down to $(x_2, f(x_2))$ and somewhere in between would have a local maximum point. (This is not obvious; it is a result of the Extreme Value Theorem.) But at that local maximum point the derivative of f would be zero or nonexistent, yet we already know that the derivative is zero or nonexistent only at x_1, x_2, and x_3. The upshot is that one computation tells us that $(x_2, f(x_2))$ has the largest y coordinate of any point on the graph near x_2 and to the left of x_2. We can perform the same test on the right. If we find that on both sides of x_2 the values are smaller, then there must be a local maximum at $(x_2, f(x_2))$; if we find that on both sides of x_2 the values are larger, then there must be a local minimum at $(x_2, f(x_2))$; if we find one of each, then there is neither a local maximum or minimum at x_2.

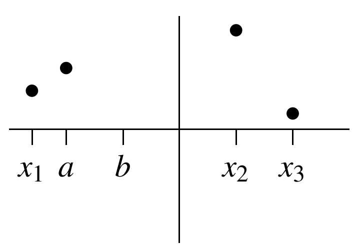


Figure 5.7: Testing for a maximum or minimum.

It is not always easy to compute the value of a function at a particular point. The task is made easier by the availability of calculators and computers, but they have their own drawbacks—they do not always allow us to distinguish between values that are very close together. Nevertheless, because this method is conceptually simple and sometimes easy to perform, you should always consider it.

Example 5.9: Simple Cubic

Find all local maximum and minimum points for the function $f(x) = x^3 - x$.

Solution. The derivative is $f'(x) = 3x^2 - 1$. This is defined everywhere and is zero at $x = \pm\sqrt{3}/3$. Looking first at $x = \sqrt{3}/3$, we see that $f(\sqrt{3}/3) = -2\sqrt{3}/9$. Now we test two points on either side of $x = \sqrt{3}/3$, choosing one point in the interval $(-\sqrt{3}/3, \sqrt{3}/3)$ and one point in the interval $(\sqrt{3}/3, \infty)$. Since $f(0) = 0 > -2\sqrt{3}/9$ and $f(1) = 0 > -2\sqrt{3}/9$, there must be a local minimum at $x = \sqrt{3}/3$. For $x = -\sqrt{3}/3$, we see that $f(-\sqrt{3}/3) = 2\sqrt{3}/9$. This time we can use $x = 0$ and $x = -1$, and we find that $f(-1) = f(0) = 0 < 2\sqrt{3}/9$, so there must be a local maximum at $x = -\sqrt{3}/3$. ♣

Of course this example is made very simple by our choice of points to test, namely $x = -1$, 0, 1. We could have used other values, say $-5/4$, $1/3$, and $3/4$, but this would have made the calculations considerably more tedious, and we should always choose very simple points to test if we can.

Example 5.10: Max and Min

Find all local maximum and minimum points for $f(x) = \sin x + \cos x$.

Solution. The derivative is $f'(x) = \cos x - \sin x$. This is always defined and is zero whenever $\cos x = \sin x$. Recalling that the $\cos x$ and $\sin x$ are the x and y coordinates of points on a unit circle, we see that $\cos x = \sin x$ when x is $\pi/4$, $\pi/4 \pm \pi$, $\pi/4 \pm 2\pi$, $\pi/4 \pm 3\pi$, etc. Since both sine and cosine have a period of 2π, we need only determine the status of $x = \pi/4$ and $x = 5\pi/4$. We can use 0 and $\pi/2$ to test the critical value $x = \pi/4$. We find that $f(\pi/4) = \sqrt{2}$, $f(0) = 1 < \sqrt{2}$ and $f(\pi/2) = 1$, so there is a local maximum when $x = \pi/4$ and also when $x = \pi/4 \pm 2\pi$, $\pi/4 \pm 4\pi$, etc. We can summarize this more neatly by saying that there are local maxima at $\pi/4 \pm 2k\pi$ for every integer k.

We use π and 2π to test the critical value $x = 5\pi/4$. The relevant values are $f(5\pi/4) = -\sqrt{2}$, $f(\pi) = -1 > -\sqrt{2}$, $f(2\pi) = 1 > -\sqrt{2}$, so there is a local minimum at $x = 5\pi/4$, $5\pi/4 \pm 2\pi$, $5\pi/4 \pm 4\pi$, etc. More succinctly, there are local minima at $5\pi/4 \pm 2k\pi$ for every integer k. ♣

Example 5.11: Max and Min

Find all local maximum and minimum points for $g(x) = x^{2/3}$.

Solution. The derivative is $g'(x) = \frac{2}{3}x^{-1/3}$. This is undefined when $x = 0$ and is not equal to zero for any x in the domain of g'. Now we test two points on either side of $x = 0$. We use $x = -1$ and $x = 1$. Since $g(0) = 0$, $g(-1) = 1 > 0$ and $g(1) = 1 > 0$, there must be a local minimum at $x = 0$. ♣

Exercises for 5.2.1

Find all local maximum and minimum points (x,y) by the method of this section.

Exercise 5.2.1 $y = x^2 - x$

Exercise 5.2.2 $y = 2 + 3x - x^3$

Exercise 5.2.3 $y = x^3 - 9x^2 + 24x$

Exercise 5.2.4 $y = x^4 - 2x^2 + 3$

Exercise 5.2.5 $y = 3x^4 - 4x^3$

Exercise 5.2.6 $y = (x^2 - 1)/x$

Exercise 5.2.7 $y = 3x^2 - (1/x^2)$

Exercise 5.2.8 $y = \cos(2x) - x$

Exercise 5.2.9 $f(x) = x^2 - 98x + 4$

Exercise 5.2.10 *For any real number x there is a unique integer n such that $n \leq x < n+1$, and the greatest integer function is defined as $\lfloor x \rfloor = n$. Where are the critical values of the greatest integer function? Which are local maxima and which are local minima?*

Exercise 5.2.11 *Explain why the function $f(x) = 1/x$ has no local maxima or minima.*

Exercise 5.2.12 *How many critical points can a quadratic polynomial function have?*

Exercise 5.2.13 *Show that a cubic polynomial can have at most two critical points. Give examples to show that a cubic polynomial can have zero, one, or two critical points.*

Exercise 5.2.14 *Explore the family of functions $f(x) = x^3 + cx + 1$ where c is a constant. How many and what types of local extremes are there? Your answer should depend on the value of c, that is, different values of c will give different answers.*

Exercise 5.2.15 *We generalize the preceding two questions. Let n be a positive integer and let f be a polynomial of degree n. How many critical points can f have? (Hint: Recall the* **Fundamental Theorem of Algebra**, *which says that a polynomial of degree n has at most n roots.)*

5.2.2. Absolute Extrema

Absolute extrema are also commonly referred to as **global extrema**. Unlike local extrema, which are only “extreme" relative to points “close to” them, an absolute (or global) extrema is “extreme" relative to *all* other points in the interval under consideration.

Definition 5.12: Absolute Maxima and Minima

A real-valued function f has an **absolute maximum** *on an interval I at x_0 if $f(x_0)$ is the largest value of f on I; in other words, $f(x_0) \geq f(x)$ for all x in the domain of f that are in I.*

A real-valued function f has an **absolute minimum** *on an interval I at x_0 if $f(x_0)$ is the smallest value of f on I; in other words, $f(x_0) \leq f(x)$ for all x in the domain of f that are in I.*

Example 5.13: Absolute Extrema

Consider the function $f(x) = x^2$ on the interval $(-\infty, \infty)$. This parabola has an absolute minimum at $x = 0$. However, it does not have an absolute maximum.

Consider the function $f(x) = |x|$ on the interval $[-1, 2]$. This graph looks like a check mark. It has an absolute minimum at $x = 0$ and an absolute maximum at $x = 2$.

Consider the function $f(x) = \cos x$ on the interval $[0, \pi]$. It has an absolute minimum at $x = \pi$ and an absolute maximum at $x = 0$.

Consider the function $f(x) = e^x$ on any interval $[a, b]$, where $a < b$. Since this exponential function is increasing, it has an absolute minimum at $x = a$ and an absolute maximum at $x = b$.

Like Fermat’s Theorem, the following theorem has an intuitive appeal. However, unlike Fermat’s Theorem, the proof relies on a more advanced concept called **compactness**, which will only be covered in a course typically entitled Analysis. So, we will be content with understanding the statement of the theorem.

Theorem 5.14: Extreme-Value Theorem

If a function f is continuous on a closed interval $[a, b]$, then f has both an absolute maximum and an absolute minimum on $[a, b]$.

Although this theorem tells us that an absolute extremum exists, it does not tell us what it is or how to find it.

Note that if an absolute extremum is inside the interval (i.e. not an endpoint), then it must also be a local extremum. This immediately tells us that to find the absolute extrema of a function on an interval, we need only examine the local extrema inside the interval, and the endpoints of the interval.

We can devise a method for finding absolute extrema for a function f on a closed interval $[a, b]$:

1. Verify the function is continuous on $[a, b]$.

2. Find the derivative and determine all critical values of f that are in $[a,b]$.

3. Evaluate the function at the critical values found in Step 2 and the end points of the interval.

4. Identify the absolute extrema.

Why must a function be continuous on a closed interval in order to use this theorem? Consider the following example.

Example 5.15: Absolute Extrema of a $1/x$

Find any absolute extrema for $f(x) = 1/x$ on the interval $[-1,1]$.

Solution. The function f is not continuous at $x = 0$. Since $0 \in [-1,1]$, f is not continous on the closed interval:

$$\lim_{x\to 0^+} f(x) = +\infty$$
$$\lim_{x\to 0^-} f(x) = -\infty,$$

so we are *unable* to apply the Extreme-Value Theorem. Therefore, $f(x) = 1/x$ does not have an absolute maximum or an absolute minimum on $[-1,1]$. ♣

However, if we consider the same function on an interval where it is continuous, the theorem will apply. This is illustrated in the following example.

Example 5.16: Absolute Extrema of a $1/x$

Find any absolute extrema for $f(x) = 1/x$ on the interval $[1,2]$.

Solution. The function f is continous on the interval, so we can apply the Extreme-Value Theorem. We begin with taking the derivative to be $f'(x) = -1/x^2$ which has a critical value at $x = 0$, but since this critical value is not in $[1,2]$ we ignore it. The only points where an extrema can occur are the endpoints of the interval. To find the maximum or minimum we can simply evaluate the function: $f(1) = 1$ and $f(2) = 1/2$, so the absolute maximum is at $x = 1$ and the absolute minimum is at $x = 2$. ♣

Why must an interval be closed in order to use the above theorem? Recall the difference between open and closed intervals. Consider a function f on the open interval $(0,1)$. If we choose successive values of x moving closer and closer to 1, what happens? Since 1 is not included in the interval we will not attain exactly the value of 1. Suppose we reach a value of 0.9999 — is it possible to get closer to 1? Yes: There are infinitely many real numbers between 0.9999 and 1. In fact, any conceivable real number close to 1 will have infinitely many real numbers between itself and 1. Now, suppose f is decreasing on $(0,1)$: As we approach 1, f will continue to decrease, even if the difference between successive values of f is slight. Similarly if f is increasing on $(0,1)$.

Consider a few more examples:

Example 5.17: Determining Absolute Extrema

Determine the absolute extrema of $f(x) = x^3 - x^2 + 1$ *on the interval* $[-1, 2]$.

Solution. First, notice f is continuous on the closed interval $[-1, 2]$, so we're able to use Theorem 5.14 to determine the absolute extrema. The derivative is $f'(x) = 3x^2 - 2x$, and the critical values are $x = 0, 2/3$ which are both in the interval $[-1, 2]$. In order to find the absolute extrema, we must consider all critical values that lie within the interval (that is, in $(-1, 2)$) *and* the endpoints of the interval.

$$\begin{aligned} f(-1) &= (-1)^3 - (-1)^2 + 1 = -1 \\ f(0) &= (0)^3 - (0)^2 + 1 = 1 \\ f(2/3) &= (2/3)^3 - (2/3)^2 + 1 = 23/27 \\ f(2) &= (2)^3 - (2)^2 + 1 = 5 \end{aligned}$$

The absolute maximum is at (2,5) and the absolute minimum is at (-1,-1). ♣

Example 5.18: Determining Absolute Extrema

Determine the absolute extrema of $f(x) = -9/x - x + 10$ *on the interval* $[2, 6]$.

Solution. First, notice f is continuous on the closed interval $[2, 6]$, so we're able to use Theorem 5.14 to determine the absolute extrema. The function is not continuous at $x = 0$, but we can ignore this fact since 0 is not in $[2, 6]$. The derivative is $f'(x) = 9/x^2 - 1$, and the critical values are $x = \pm 3$, but only $x = +3$ is in the interval. In order to find the absolute extrema, we must consider all critical values that lie within the interval *and* the endpoints of the interval.

$$\begin{aligned} f(2) &= -9/(2) - (2) + 10 = 7/2 = 3.5 \\ f(3) &= -9/(3) - (3) + 10 = 4 \\ f(6) &= -9/(6) - (6) + 10 = 5/2 = 2.5 \end{aligned}$$

The absolute maximum is at (3,4) and the absolute minimum is at (6,2.5). ♣

When we are trying to find the absolute extrema of a function on an open interval, we cannot use the Extreme Value Theorem. However, if the function is continuous on the interval, many of the same ideas apply. In particular, if an absolute extremum exists, it must also be a local extremum. In addition to checking values at the local extrema, we must check the behaviour of the function as it approaches the ends of the interval.

Some examples to illustrate this method.

Example 5.19: Extrema of Secant

Find the extrema of $\sec(x)$ *on* $(-\pi/2, \pi/2)$.

Solution. Notice $\sec(x)$ is continuous on $(-\pi/2, \pi/2)$ and has one local minimum at 0. Also

$$\lim_{x \to (-\pi/2)^+} \sec(x) = \lim_{x \to (\pi/2)^-} \sec(x) = +\infty,$$

so $\sec(x)$ has no absolute maximum, but the point $(0,1)$ is the absolute minimum. ♣

A similar approach can be used for infinite intervals.

Example 5.20: Extrema of $\frac{x^2}{x^2+1}$

Find the extrema of $\dfrac{x^2}{x^2+1}$ *on* $(-\infty,\infty)$.

Solution. Since $x^2+1 \neq 0$ for all x in $(-\infty,\infty)$ the function is continuous on this interval. This function has only one critical value at $x=0$, which is the local minimum and also the absolute minimum. Now, $\lim_{x\to\pm\infty}\frac{x^2}{x^2+1}=1$, so the function does not have an absolute maximum: It continues to increase towards 1, but does not attain this exact value. ♣

Exercises for 5.2.2

Exercise 5.2.16 *Find the absolute extrema for* $f(x)=-\frac{x+4}{x-4}$ *on* $[0,3]$.

Exercise 5.2.17 *Find the absolute extrema for* $f(x)=-\frac{x+4}{x-4}$ *on* $[0,3]$.

Exercise 5.2.18 *Find the absolute extrema for* $f(x)=\csc(x)$ *on* $[0,\pi]$.

Exercise 5.2.19 *Find the absolute extrema for* $f(x)=\ln(x)/x^2$ *on* $[1,4]$.

Exercise 5.2.20 *Find the absolute extrema for* $f(x)=x\sqrt{1-x^2}$ *on* $[-1,1]$.

Exercise 5.2.21 *Find the absolute extrema for* $f(x)=xe^{-x^2/32}$ *on* $[0,2]$.

Exercise 5.2.22 *Find the absolute extrema for* $f(x)=x-\tan^{-1}(2x)$ *on* $[0,2]$.

Exercise 5.2.23 *Find the absolute extrema for* $f(x)=\frac{x}{x^2+1}$.

Exercise 5.2.24 *For each of the following, sketch a potential graph of a continuous function on the closed interval* $[0,4]$ *with the given properties.*

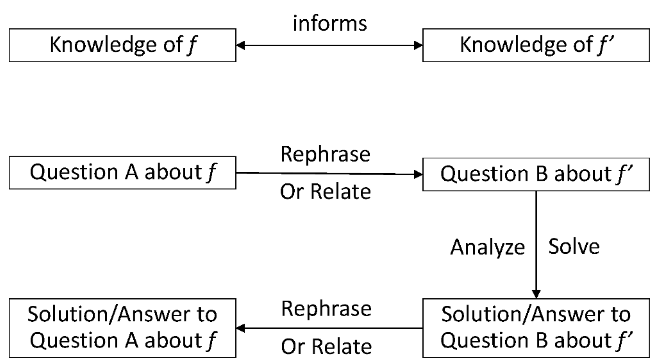
(a) *Absolute minimum at 0, absolute maximum at 2, local minimum at 3.*

(b) *Absolute maximum at 1, absolute minimum at 2, local maximum at 3.*

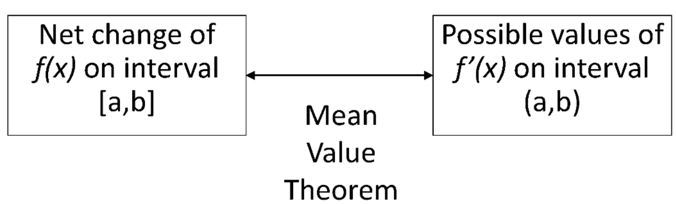
(c) *Absolute minimum at 4, absolute maximum at 1, local minimum at 2, local maxima at 1 and 3.*

5.3 The Mean Value Theorem

There are numerous applications of the derivative through its **definition** as rate of change and as the slope of the tangent line. In this section we shall look at some deeper reasons why the derivative turns out to be so useful. The simple answer is that **the derivative of a function tells us a lot about the function**. More important, "hard" questions about a function can sometimes be answered by solving a relatively simple problem about the derivative of the function.



The Mean Value Theorem tells us that there is an intimate connection between the net change of the value of any "sufficiently nice" function over an interval and the possible values of its derivative on that interval. Because of this connection, we can draw conclusions about the possible values of the derivative based on information about the values of the function, and conversely, we can draw conclusions about the values of the function based on information about the values of its derivative.



Let us illustrate the idea through the following two interesting questions involving derivatives:

1. Suppose two different functions have the same derivative; what can you say about the relationship between the two functions?

2. Suppose you drive a car from toll booth on a toll road to another toll booth at an average speed of 70 miles per hour. What can be concluded about your actual speed during the trip? In particular, did you exceed the 65 mile per hour speed limit?

While these sound very different, it turns out that the two problems are very closely related. We know that "speed" is really the derivative by a different name; let's start by translating the second question into something that may be easier to visualize. Suppose that the function $f(t)$ gives the position of your car on the toll road at time t. Your change in position between one toll booth and the next is given by $f(t_1) - f(t_0)$, assuming that at time t_0 you were at the first booth and at time t_1 you arrived at the second booth. Your average speed for the trip is $(f(t_1) - f(t_0))/(t_1 - t_0)$. If we think about the graph of $f(t)$, the average speed is the slope of the line that connects the two points $(t_0, f(t_0))$ and $(t_1, f(t_1))$. Your speed at any particular time t between t_0 and t_1 is $f'(t)$, the slope of the curve. Now question (2) becomes a question about slope. In particular, if the slope between endpoints is 70, what can be said of the slopes at points between the endpoints?

As a general rule, when faced with a new problem it is often a good idea to examine one or more simplified versions of the problem, in the hope that this will lead to an understanding of the original problem. In this case, the problem in its "slope" form is somewhat easier to simplify than the original, but equivalent, problem.

Here is a special instance of the problem. Suppose that $f(t_0) = f(t_1)$. Then the two endpoints have the same height and the slope of the line connecting the endpoints is zero. What can we say about the slope between the endpoints? It shouldn't take much experimentation before you are convinced of the truth of this statement: Somewhere between t_0 and t_1 the slope is exactly zero, that is, somewhere between t_0 and t_1 the slope is equal to the slope of the line between the endpoints. This suggests that perhaps the same is true even if the endpoints are at different heights, and again a bit of experimentation will probably convince you that this is so. But we can do better than "experimentation"—we can prove that this is so.

We start with the simplified version:

Theorem 5.21: Rolle's Theorem

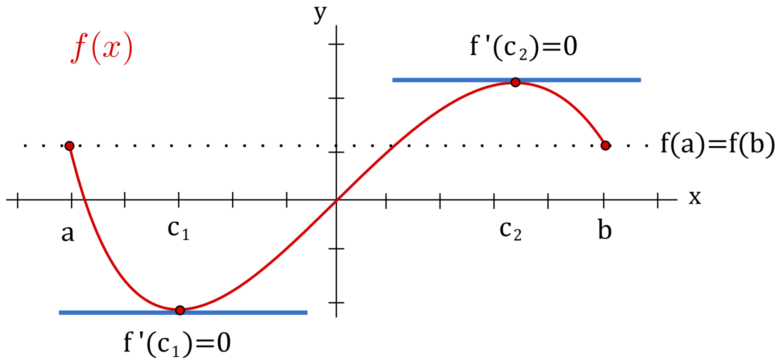
Suppose that $f(x)$ has a derivative on the interval (a,b), is continuous on the interval $[a,b]$, and $f(a) = f(b)$. Then at some value $c \in (a,b)$, $f'(c) = 0$.

Proof. We know that $f(x)$ has a maximum and minimum value on $[a,b]$ (because it is continuous), and we also know that the maximum and minimum must occur at an endpoint, at a point at which the derivative is zero, or at a point where the derivative is undefined. Since the derivative is never undefined, that possibility is removed.

If the maximum or minimum occurs at a point c, other than an endpoint, where $f'(c) = 0$, then we have found the point we seek. Otherwise, the maximum and minimum both occur at an endpoint, and since the endpoints have the same height, the maximum and minimum are the same. This means that $f(x) = f(a) = f(b)$ at every $x \in [a,b]$, so the function is a horizontal line, and it has derivative zero everywhere in (a,b). Then we may choose any c at all to get $f'(c) = 0$. ♣

Rolle's Theorem is illustrated below for a function $f(x)$ where $f'(x) = 0$ holds for two values of $x = c_1$

and $x = c_2$:



Perhaps remarkably, this special case is all we need to prove the more general one as well.

Theorem 5.22: Mean Value Theorem

Suppose that $f(x)$ has a derivative on the interval (a,b) and is continuous on the interval $[a,b]$. Then at some value $c \in (a,b)$, $f'(c) = \dfrac{f(b)-f(a)}{b-a}$.

Proof. Let $m = \dfrac{f(b)-f(a)}{b-a}$, and consider a new function $g(x) = f(x) - m(x-a) - f(a)$. We know that $g(x)$ has a derivative everywhere, since $g'(x) = f'(x) - m$. We can compute $g(a) = f(a) - m(a-a) - f(a) = 0$ and

$$g(b) = f(b) - m(b-a) - f(a) \quad = f(b) - \tfrac{f(b)-f(a)}{b-a}(b-a) - f(a)$$
$$= f(b) - (f(b) - f(a)) - f(a) = 0.$$

So the height of $g(x)$ is the same at both endpoints. This means, by Rolle's Theorem, that at some c, $g'(c) = 0$. But we know that $g'(c) = f'(c) - m$, so

$$0 = f'(c) - m = f'(c) - \frac{f(b)-f(a)}{b-a},$$

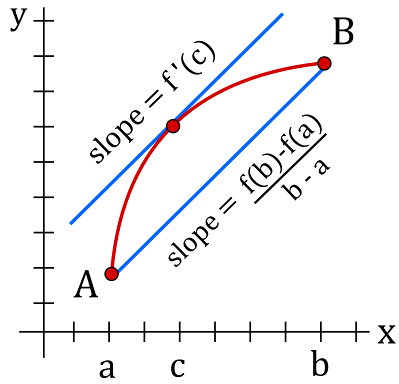
which turns into

$$f'(c) = \frac{f(b)-f(a)}{b-a},$$

exactly what we want. ♣

The Mean Value Theorem is illustrated below showing the existence of a point $x = c$ for a function $f(x)$ where the tangent line at $x = c$ (with slope $f'(c)$) is parallel to the secant line connecting $A(a, f(a))$

and $B(b, f(b))$ (with slope $\frac{f(b)-f(a)}{b-a}$):



Returning to the original formulation of question (2), we see that if $f(t)$ gives the position of your car at time t, then the Mean Value Theorem says that at some time c, $f'(c) = 70$, that is, at some time you must have been traveling at exactly your average speed for the trip, and that indeed you exceeded the speed limit.

Now let's return to question (1). Suppose, for example, that two functions are known to have derivative equal to 5 everywhere, $f'(x) = g'(x) = 5$. It is easy to find such functions: $5x$, $5x+47$, $5x-132$, etc. Are there other, more complicated, examples? No—the only functions that work are the "obvious" ones, namely, $5x$ plus some constant. How can we see that this is true?

Although "5" is a very simple derivative, let's look at an even simpler one. Suppose that $f'(x) = g'(x) = 0$. Again we can find examples: $f(x) = 0$, $f(x) = 47$, $f(x) = -511$ all have $f'(x) = 0$. Are there non-constant functions f with derivative 0? No, and here's why: Suppose that $f(x)$ is not a constant function. This means that there are two points on the function with different heights, say $f(a) \neq f(b)$. The Mean Value Theorem tells us that at some point c, $f'(c) = (f(b) - f(a))/(b-a) \neq 0$. So any non-constant function does not have a derivative that is zero everywhere; this is the same as saying that the only functions with zero derivative are the constant functions.

Let's go back to the slightly less easy example: suppose that $f'(x) = g'(x) = 5$. Then $(f(x) - g(x))' = f'(x) - g'(x) = 5 - 5 = 0$. So using what we discovered in the previous paragraph, we know that $f(x) - g(x) = k$, for some constant k. So any two functions with derivative 5 must differ by a constant; since $5x$ is known to work, the only other examples must look like $5x + k$.

Now we can extend this to more complicated functions, without any extra work. Suppose that $f'(x) = g'(x)$. Then as before $(f(x) - g(x))' = f'(x) - g'(x) = 0$, so $f(x) - g(x) = k$. Again this means that if we find just a single function $g(x)$ with a certain derivative, then every other function with the same derivative must be of the form $g(x) + k$.

Example 5.23: Given Derivative

Describe all functions that have derivative $5x-3$.

Solution. It's easy to find one: $g(x) = (5/2)x^2 - 3x$ has $g'(x) = 5x - 3$. The only other functions with the same derivative are therefore of the form $f(x) = (5/2)x^2 - 3x + k$.

Alternately, though not obviously, you might have first noticed that $g(x) = (5/2)x^2 - 3x + 47$ has $g'(x) = 5x - 3$. Then every other function with the same derivative must have the form $f(x) = (5/2)x^2 -$

$3x+47+k$. This looks different, but it really isn't. The functions of the form $f(x)=(5/2)x^2-3x+k$ are exactly the same as the ones of the form $f(x)=(5/2)x^2-3x+47+k$. For example, $(5/2)x^2-3x+10$ is the same as $(5/2)x^2-3x+47+(-37)$, and the first is of the first form while the second has the second form. ♣

This is worth calling a theorem:

Theorem 5.24: Functions with the Same Derivative

If $f'(x)=g'(x)$ for every $x\in(a,b)$, then for some constant k, $f(x)=g(x)+k$ on the interval (a,b).

Example 5.25: Same Derivative

Describe all functions with derivative $\sin x+e^x$. One such function is $-\cos x+e^x$, so all such functions have the form $-\cos x+e^x+k$.

Theorem 5.24 and the above example illustrate what the Mean Value Theorem allows us to say about $f(x)$ when we have perfect information about $f'(x)$. Specifically, $f(x)$ is determined up to a constant. Our next example illustrates almost the opposite extreme situation, one where we have much less information about $f'(x)$ beyond the fact that $f'(x)$ exists. Specifically, assuming that we know an upper bound on the values of $f'(x)$, what can we say about the values of $f(x)$?

Example 5.26: Conclusion Regarding Function Value Based on Derivative Information

Suppose that f is a differentiable function such that $f'(x)\le 2$ for all x. What is the largest possible value of $f(7)$ if $f(3)=5$?

Solution. We are interested in the values of $f(x)$ at $x=3$ and $x=7$. It makes sense to focus our attention on the interval between 3 and 7. It is given that $f(x)$ is differentiable for all x. So, $f(x)$ is also continuous at all x. In particular, $f(x)$ is continuous on the interval $[3,7]$ and differentiable on the interval $(3,7)$. By the Mean Value Theorem, we know that there is some c in $(3,7)$ such that

$$f'(c)=\frac{f(7)-f(3)}{7-3}.$$

Simplifying and using the given information $f(3)=5$, we get

$$f'(c)=\frac{f(7)-5}{4},$$

or, after re-arranging the terms,

$$f(7)=4f'(c)+5.$$

We do not know the exact value of c, but we do know that $f'(x)\le 2$ for all x. This implies that $f'(c)\le 2$. Therefore,

$$f(7)\le 4\cdot 2+5=13.$$

That is, the value of $f(7)$ cannot exceed 13. To convince ourselves that 13 (as opposed to some smaller number) is the largest possible value of $f(7)$, we still need to show that it is possible for the value of $f(7)$ to reach 13. If we review our proof, we notice that the inequality will be an equality if $f'(c) = 2$. One way to guarantee this without knowing anything about c is to require $f'(x) = 2$ for all x. This means that $f(x) = 2x + k$ for some constant k. From the condition $f(3) = 5$, we see that $k = -1$. We can easily verify that indeed $f(x) = 2x - 1$ meets all our requirements and $f(7) = 13$. ♣

Exercises for Section 5.3

Exercise 5.3.1 *Let $f(x) = x^2$. Find a value $c \in (-1,2)$ so that $f'(c)$ equals the slope between the endpoints of $f(x)$ on $[-1,2]$.*

Exercise 5.3.2 *Verify that $f(x) = x/(x+2)$ satisfies the hypotheses of the Mean Value Theorem on the interval $[1,4]$ and then find all of the values, c, that satisfy the conclusion of the theorem.*

Exercise 5.3.3 *Verify that $f(x) = 3x/(x+7)$ satisfies the hypotheses of the Mean Value Theorem on the interval $[-2,6]$ and then find all of the values, c, that satisfy the conclusion of the theorem.*

Exercise 5.3.4 *Let $f(x) = \tan x$. Show that $f(\pi) = f(2\pi) = 0$ but there is no number $c \in (\pi, 2\pi)$ such that $f'(c) = 0$. Why does this not contradict Rolle's theorem?*

Exercise 5.3.5 *Let $f(x) = (x-3)^{-2}$. Show that there is no value $c \in (1,4)$ such that $f'(c) = (f(4) - f(1))/(4-1)$. Why is this not a contradiction of the Mean Value Theorem?*

Exercise 5.3.6 *Describe all functions with derivative $x^2 + 47x - 5$.*

Exercise 5.3.7 *Describe all functions with derivative $\dfrac{1}{1+x^2}$.*

Exercise 5.3.8 *Describe all functions with derivative $x^3 - \dfrac{1}{x}$.*

Exercise 5.3.9 *Describe all functions with derivative $\sin(2x)$.*

Exercise 5.3.10 *Find $f(x)$ if $f'(x) = e^{-x}$ and $f(0) = 2$.*

Exercise 5.3.11 *Suppose that f is a differentiable function such that $f'(x) \geq -3$ for all x. What is the smallest possible value of $f(4)$ if $f(-1) = 2$?*

Exercise 5.3.12 *Show that the equation $6x^4 - 7x + 1 = 0$ does not have more than two distinct real roots.*

Exercise 5.3.13 *Let f be differentiable on $\mathbb{R}$. Suppose that $f'(x) \neq 0$ for every x. Prove that f has at most one real root.*

Exercise 5.3.14 *Prove that for all real x and y $|\cos x - \cos y| \le |x - y|$. State and prove an analogous result involving sine.*

Exercise 5.3.15 *Show that $\sqrt{1+x} \le 1 + (x/2)$ if $-1 < x < 1$.*

Exercise 5.3.16 *Suppose that $f(a) = g(a)$ and that $f'(x) \le g'(x)$ for all $x \ge a$.*

(a) *Prove that $f(x) \le g(x)$ for all $x \ge a$.*

(b) *Use part (a) to prove that $e^x \ge 1 + x$ for all $x \ge 0$.*

(c) *Use parts (a) and (b) to prove that $e^x \ge 1 + x + \dfrac{x^2}{2}$ for all $x \ge 0$.*

(d) *Can you generalize these results?*

5.4 Linear and Higher Order Approximations

When we define the derivative $f'(x)$ as the rate of change of $f(x)$ with respect to x, we notice that in relation to the graph of f, the derivative is the slope of the tangent line, which (loosely speaking) is the line that just grazes the graph. But what precisely do we mean by this? In short, **the tangent line approximates the graph near the point of contact**. The definition of the derivative $f'(a)$ guarantees this when it exists: By taking x sufficiently close to a but not equal to a,

$$\frac{f(x) - f(a)}{x - a} \approx f'(a),$$

and consequently,

$$f(x) \approx f'(a)(x-a) + f(a).$$

The left hand side gives us the y-value of the function $y = f(x)$ and the right hand side gives us the y-value $y = f'(a)(x-a) + f(a)$ for the tangent line to the graph of f at the point $(a, f(a))$.

In this section we will explore how to apply this idea to approximate some values of f, some changes in the values of f, and also the roots of f.

5.4.1. Linear Approximations

We begin by the first derivative as an application of the tangent line to approximate f.

Recall that the tangent line to $f(x)$ at a point $x = a$ is given by

$$L(x) = f'(a)(x-a) + f(a).$$

The tangent line in this context is also called the **linear approximation** to f at a.

If f is differentiable at a then L is a good approximation of f so long as x is "not too far" from a. Put another way, if f is differentiable at a then under a microscope f will look very much like a straight line, and thus will look very much like L; since $L(x)$ is often much easier to compute than $f(x)$, then it makes sense to use L as an approximation. Figure 5.8 shows a tangent line to $y = x^2$ at three different magnifications.

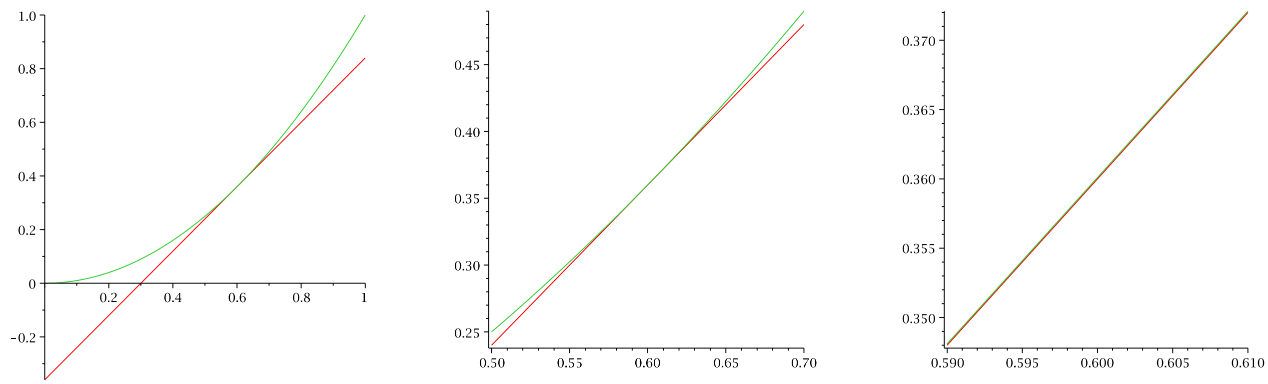


Figure 5.8: The linear approximation to $y = x^2$.

Thus in practice if we want to approximate a difficult value of $f(b)$, then we may be able to approximate this value using a linear approximation, provided that we can compute the tangent line at some point a close to b. Here is an example.

Example 5.27: Linear Approximation

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

Solution. We are asked to calculate $f(6) = \sqrt{6+4} = \sqrt{10}$ which is not easy to do without a calculator. However 9 is (relatively) close to 10 and of course $f(5) = \sqrt{9}$ is easy to compute, and we use this to approximate $\sqrt{10}$.

To do so we have $f'(x) = 1/(2\sqrt{x+4})$, and thus the linear approximation to f at $x = 5$ is

$$L(x) = \left(\frac{1}{2\sqrt{5+4}}\right)(x-5) + \sqrt{5+4} = \frac{x-5}{6} + 3.$$

Now to estimate $\sqrt{10}$, we substitute 6 into the linear approximation $L(x)$ instead of $f(x)$, to obtain

$$\sqrt{6+4} \approx \frac{6-5}{6} + 3 = \frac{19}{6} = 3{}^1\!/\!{}_6 = 3.1\bar{6} \approx 3.17$$

It turns out the exact value of $\sqrt{10}$ is actually $3.16227766\ldots$ but our estimate of 3.17 was very easy to obtain and is relatively accurate. This estimate is only accurate to one decimal place. ♣

With modern calculators and computing software it may not appear necessary to use linear approximations, but in fact they are quite useful. For example in cases requiring an explicit numerical approximation, they allow us to get a quick estimate which can be used as a "reality check" on a more complex calculation. Further in some complex calculations involving functions, the linear approximation makes an otherwise intractable calculation possible without serious loss of accuracy.

Example 5.28: Linear Approximation of Sine

Find the linear approximation of $\sin x$ *at* $x = 0$*, and use it to compute small values of* $\sin x$*.*

Solution. If $f(x) = \sin x$, then $f'(x) = \cos x$, and thus the linear approximation of $\sin x$ at $x = 0$ is:

$$L(x) = \cos(0)(x-0) + \sin(0) = x.$$

Thus when x is small this is quite a good approximation and is used frequently by engineers and scientists to simplify some calculations.

For example you can use your calculator (in radian mode since the derivative of $\sin x$ is $\cos x$ only in radian) to see that

$$\sin(0.1) = 0.099833416\ldots$$

and thus $L(0.1) = 0.1$ is a very good and quick approximation without any calculator! ♣

Exercises for 5.4.1

Exercise 5.4.1 *Find the linearization* $L(x)$ *of* $f(x) = \ln(1+x)$ *at* $a = 0$*. Use this linearization to approximate* $f(0.1)$*.*

Exercise 5.4.2 *Use linear approximation to estimate* $(1.9)^3$*.*

Exercise 5.4.3 *Show in detail that the linear approximation of* $\sin x$ *at* $x = 0$ *is* $L(x) = x$ *and the linear approximation of* $\cos x$ *at* $x = 0$ *is* $L(x) = 1$*.*

Exercise 5.4.4 *Use* $f(x) = \sqrt[3]{x+1}$ *to approximate* $\sqrt[3]{9}$ *by choosing an appropriate point* $x = a$*. Are we over- or under-estimating the value of* $\sqrt[3]{9}$*? Explain.*

5.4.2. Differentials

Very much related to linear approximations are the *differentials* dx and dy, used not to approximate values of f, but instead the change (or rise) in the values of f.

Definition 5.29: Differentials dx and dy

Let $y = f(x)$ *be a differentiable function. We define a new independent variable* dx*, and a new dependent variable* $dy = f'(x)\,dx$*. Notice that* dy *is a function both of* x *(since* $f'(x)$ *is a function of* x*) and of* dx*. We call both* dx *and* dy **differentials**.

Now fix a point a and let $\Delta x = x - a$ and $\Delta y = f(x) - f(a)$. If x is near a then Δx is clearly small. If we set $dx = \Delta x$ then we obtain

$$dy = f'(a)\,dx \approx \frac{\Delta y}{\Delta x}\Delta x = \Delta y.$$

Thus, dy can be used to approximate Δy, the actual change in the function f between a and x. This is exactly the approximation given by the tangent line:

$$dy = f'(a)(x-a) = f'(a)(x-a) + f(a) - f(a) = L(x) - f(a).$$

While $L(x)$ approximates $f(x)$, dy approximates how $f(x)$ has changed from $f(a)$. Figure 5.9 illustrates the relationships.

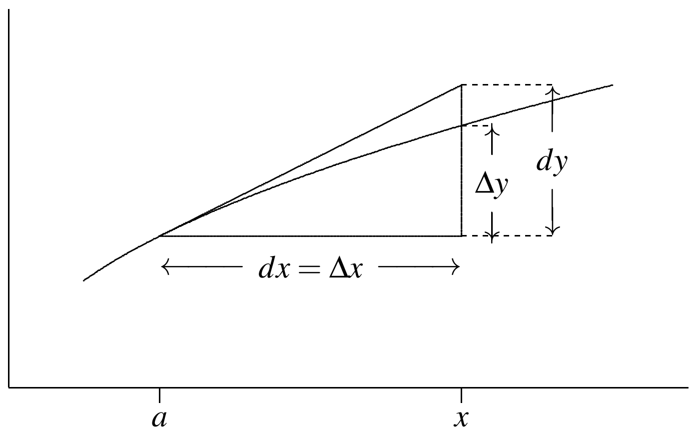


Figure 5.9: Differentials.

Here is a concrete example.

Example 5.30: Rise of Natural Logarithm

Approximate the rise of $f(x) = \ln x$ from $x = 1$ to $x = 1.1$, using linear approximation.

Solution. Note that $\ln(1.1)$ is not readily calculated (without a calculator) hence why we wish to use linear approximation to approximate $f(1.1) - f(1)$.

We fix $a = 1$ and as above we have $\Delta x = x - 1$ and $\Delta y = f(x) - f(1) = \ln x$, and obtain

$$dy = f'(1)dx \approx \frac{\Delta y}{\Delta x}\Delta x = \Delta y.$$

But $f'(x) = 1/x$ and thus $f'(1) = 1/1 = 1$, we obtain in this case

$$dy = dx \approx \Delta y.$$

Finally for $x = 1.1$, we can easily approximate the rise of f as

$$f(1.1) - f(1) = \Delta y \approx dy = 1.1 - 1 = 0.1.$$

The correct value of $\ln(1.1) = \ln 1$ is $0.0953\ldots$ and thus we were relatively close. ♣

Exercises for 5.4.2

Exercise 5.4.5 *Let* $f(x) = x^4$. *If* $a = 1$ *and* $dx = \Delta x = 1/2$, *what are* Δy *and* dy?

Exercise 5.4.6 *Let* $f(x) = \sqrt{x}$. *If* $a = 1$ *and* $dx = \Delta x = 1/10$, *what are* Δy *and* dy?

Exercise 5.4.7 *Let* $f(x) = \sin(2x)$. *If* $a = \pi$ *and* $dx = \Delta x = \pi/100$, *what are* Δy *and* dy?

Exercise 5.4.8 *Use differentials to estimate the amount of paint needed to apply a coat of paint 0.02 cm thick to a sphere with diameter* 40 *meters. (Recall that the volume of a sphere of radius r is* $V = (4/3)\pi r^3$. *Notice that you are given that* $dr = 0.02$.*)*

5.4.3. Taylor Polynomials

We can go beyond first order derivatives to create polynomials approximating a function as closely as we wish, these are called *Taylor Polynomials*.

While our linear approximation $L(x) = f'(a)(x-a) + f(a)$ at a point a was a polynomial of degree 1 such that both $L(a) = f(a)$ and $L'(a) = f'(a)$, we can now form a polynomial

$$T_n(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + \cdots + a_n(x-a)^n$$

which has the same first n derivatives at $x = a$ as the function f.

By successively computing the derivatives of T_n, we obtain:

$$\begin{aligned} a_0 &= f(a) = \tfrac{f(a)}{0!} \\ a_1 &= \tfrac{f'(a)}{1!} \\ a_2 &= \tfrac{f''(a)}{2!} \\ \cdots \\ a_k &= \tfrac{f^{(k)}(a)}{k!} \\ \cdots a_n &= \tfrac{f^{(n)}(a)}{n!} \end{aligned}$$

where $f^{(k)}(x)$ is the k^{th} derivative of $f(x)$, and $n! = n(n-1)(n-2)\ldots(2)(1)$, referred to as *factorial* notation.

Here is an example.

Example 5.31: Approximate e using Taylor Polynomials

Approximate e^x *using Taylor polynomials at* $a = 0$, *and use this to approximate* e.

Solution. In this case we use the function $f(x) = e^x$ at $a = 0$, and therefore

$$T_n(x) = a_0 + a_1x + a_xx^2 + a_3x^3 + \ldots + a_nx^n$$

Since all derivatives $f^{(k)}(x) = e^x$, we get:

$$\begin{aligned}
&a_0 = f(0) = 1\\
&a_1 = \tfrac{f'(0)}{1!} = 1\\
&a_2 = \tfrac{f''(0)}{2!} = \tfrac{1}{2!}\\
&a_3 = \tfrac{f'''(0)}{3!} = \tfrac{1}{3!}\\
&\ldots\\
&a_k = \tfrac{f^{(k)}(0)}{k!} = \tfrac{1}{k!}\\
&\ldots\\
&a_n = \tfrac{f^{(n)}(0)}{n!} = \tfrac{1}{n!}
\end{aligned}$$

Thus

$$\begin{aligned}
&T_1(x) = 1 + x = L(x)\\
&T_2(x) = 1 + x + \tfrac{x^2}{2!}\\
&T_3(x) = 1 + x + \tfrac{x^2}{2!} + \tfrac{x^3}{3!}
\end{aligned}$$

and in general

$$T_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}.$$

Finally we can approximate $e = f(1)$ by simply calculating $T_n(1)$. A few values are:

$$\begin{aligned}
&T_1(1) = 1 + 1 = 2\\
&T_2(1) = 1 + 1 + \tfrac{1^2}{2!} = 2.5\\
&T_4(1) = 1 + 1 + \tfrac{1^2}{2!} + \tfrac{1^3}{3!} = 2.\overline{6}\\
&T_8(1) = 2.71825396825\\
&T_{20}(1) = 2.71828182845
\end{aligned}$$

We can continue this way for larger values of n, but $T_{20}(1)$ is already a pretty good approximation of e, and we took only 20 terms! ♣

Exercises for 5.4.3

Exercise 5.4.9 *Find the* 5^{th} *degree Taylor polynomial for* $f(x) = \sin x$ *around* $a = 0$.

(a) Use this Taylor polynomial to approximate $\sin(0.1)$.

(b) Use a calculator to find $\sin(0.1)$. *How does this compare to our approximation in part (a)?*

Exercise 5.4.10 *Suppose that* f'' *exists and is continuous on* $[1,2]$. *Suppose also that* $|f''(x)| \leq \frac{1}{4}$ *for all* x *in* $(1,2)$. *Prove that if we use the linearization* $y = L(x)$ *of* $y = f(x)$ *at* $x = 1$ *as an approximation of* $y = f(x)$ *near* $x = 1$, *then our estimated value of* $f(1.2)$ *is guaranteed to have an accuracy of at least 0.01, i.e., our estimate will lie within 0.01 units of the true value.*

Exercise 5.4.11 *Find the 3rd degree Taylor polynomial for* $f(x) = \frac{1}{1-x} - 1$ *around* $a = 0$*. Explain why this approximation would not be useful for calculating* $f(5)$*.*

Exercise 5.4.12 *Consider* $f(x) = \ln x$ *around* $a = 1$*.*

(a) *Find a general formula for* $f^{(n)}(x)$ *for* $n \geq 1$*.*

(b) *Find a general formula for the Taylor Polynomial,* $T_n(x)$*.*

5.4.4. Newton's Method

A well known numeric method is *Newton's Method* (also sometimes referred to as *Newton-Raphson's Method*), named after Isaac Newton and Joseph Raphson. This method is used to find roots, or x-intercepts, of a function. While we may be able to find the roots of a polynomial which we can easily factor, we saw in the previous chapter on **Limits**, that for example the function $e^x + x = 0$ has a solution (*i.e.* root, or x-intercept) at $x \approx -0.56714$. By the *Intermediate Value Theorem* we know that the function $e^x + x = 0$ does have a solution. We cannot here simply solve for such a root algebraically, but we can use a numerical method such as $Newton's$. Such a process is typically classified as an *iterative* method, a name given to a technique which involves repeating similar steps until the desired accuracy is obtained. Many computer algorithms are coded with a for-loop, repeating an iterative step to converge to a solution.

The idea is to start with an initial value x_0 (approximating the root), and use linear approximation to create values $x_1, x_2, \cdots$ getting closer and closer to a root.

The first value x_1 corresponds to the intercept of the tangent line of $f(x_0)$ with the x-axis, which is:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

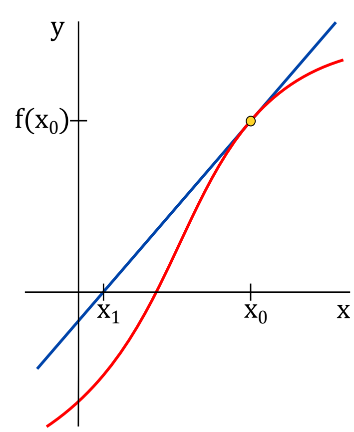


Figure 5.10: First iteration of Newton's Method.

We can see in Figure 5.10, that if we compare the point $(x_0, 0)$ to $(x_1, 0)$, we would likely come to the conclusion that $(x_1, 0)$ is closer to the actual root of $f(x)$ than our original guess, $(x_0, 0)$. As will be discussed, the choice of x_0 must be done correctly, and it may occur that x_1 does not yield a better estimate of the root.

Newton's method is simply to repeat this process again and again in an effort to obtain a more accurate solution. Thus at the next step we obtain:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

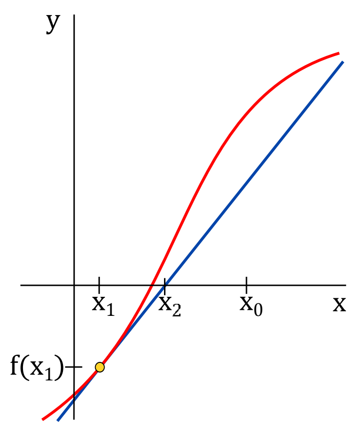


Figure 5.11: Second iteration of Newton's Method.

We can now clearly see how $(x_2,0)$ is a better estimate of the root of $f(x)$, rather than any of the previous points. Moving forward, we will get:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$$

Rest assured, $(x_3,0)$ will be an even better estimate of the root! We express the general iterative step as:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

The idea is to iterate these steps to obtain the desired accuracy. Here is an example.

Example 5.32: Newton method to approximate roots

Use Newton's method to approximate the roots of $f(x) = x^3 - x + 1$.

Solution. You can try to find solve the equation algebraically to see that this is a difficult task, and thus it make sense to try a numerical method such as Newton's.

To find an initial value x_0, note that $f(-1) = -5$ and $f(0) = 1$, and by the Intermediate Value Theorem this f has a root between these two values, and we decide to start with $x_0 = -1$ (you can try other values to see what happens).

Note that $f'(x) = 3x^2 - 1$, and thus we get

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x + n - \frac{x_n^3 - x_n + 1}{3x_n^2 - 1}$$

Thus we can produce the following values (try it):

$$\begin{aligned} x_0 &= -1 \\ x_1 &= -1.5000 \\ x_2 &= -1.347826.. \\ x_3 &= -1.325200.. \\ x_4 &= -1.324718.. \\ x_5 &= -1.324717.. \\ x_6 &= -1.324717.. \\ &\cdots \end{aligned}$$

and we can now approximate the root as -1.324717. ♣

As with any numerical method, we need to be aware of the weaknesses of any technique we are using.

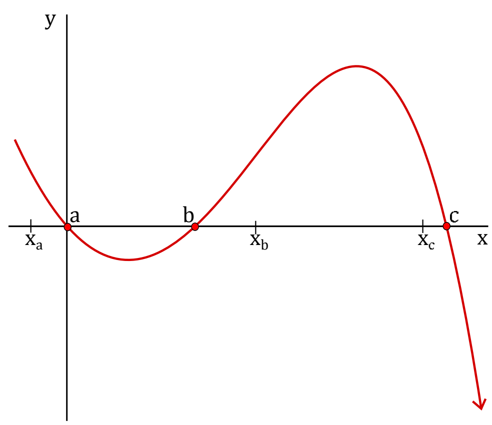


Figure 5.12: Function with three distinct solutions.

If we know our root is somewhere near a, we would make our guess $x_0 = a$. Generally speaking, a good practice is to make our guess as close to the actual root as possible. In some cases we may have no idea where the root is, so it would be prudent to perform the algorithm several times on several different initial guesses and analyze the results.

For example we can see in Figure 5.12 that $f(x)$ in fact has three roots, and depending on our initial guess, we may get the algorithm to converge to different roots. If we did not know where the roots were, we would try the technique several times. In one instance, if our initial guess was x_a, we'd likely converge to $(a, 0)$. Then if we were to choose another guess, x_b, then we'd likely converge to $(b, 0)$. Eventually, using various initial guesses we'd get one of three roots: a, b, or c. Under these circumstances we can clearly see the effectiveness of this numeric method.

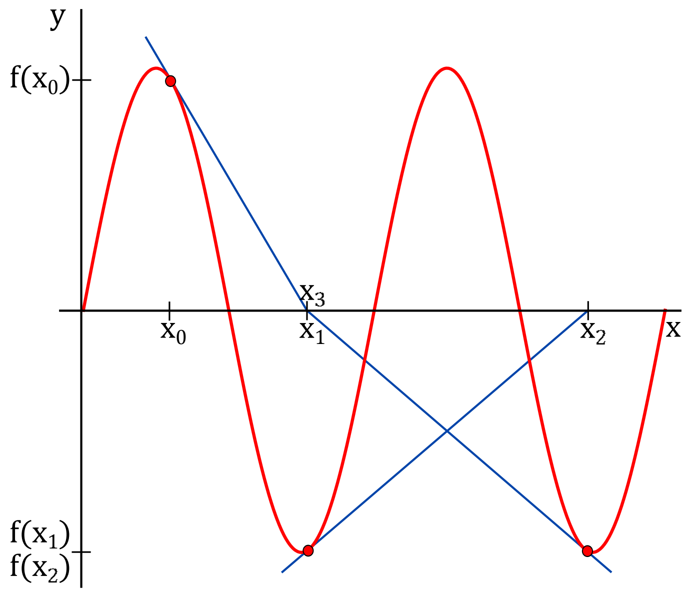


Figure 5.13: Newton's Method applied to $\sin x$**.**

As another example if we attempt to use *Newton's Method* on $f(x) = \sin x$ using $x_0 = \pi/2$, then $f'(x_0) = 0$ so x_1 is undefined and we cannot proceed. Even in general x_{n+1} is typically nowhere near x_n, and in general not converging to the root nearest to our initial guess of x_0. In effect, the algorithm keeps "bouncing around". An example of which is depicted in Figure 5.13. Based on our initial guess for such a function, the algorithm may or may not converge to a root, or it may or may not converge to the root **closest** to the initial guess. This gives rise to the more common issue: Selection of the initial guess, x_0.

Here is a summary.

Key Points in using Newton's method to approximate a root of $f(x)$

1. Choosing x_0 as close as possible to the root we wish to find.
2. A guess for x_0 which makes the algorithm "bounce around" is considered *unstable*.
3. Even the smallest changes to x_0 can have drastic effects: We may converge to another root, we may converge very slowly (requiring many more iterations), or we may encounter an unstable point.
4. We may encounter a *stationary point* if we choose x_0 such that $f'(x) = 0$ (*i.e.* at a critical point!) in which case the algorithm fails.

This is all to say that your initial guess for x_0 can be extremely important.

Exercises for 5.4.4

Exercise 5.4.13 *Use Newton's Method to find all roots of* $f(x) = 3x^2 - 9x - 11$*. (Hint: use Intermediate Value Theorem to choose an appropriate* x_0*)*

Exercise 5.4.14 *Consider* $f(x) = x^3 - x^2 + x - 1$*.*

(a) *Using initial approximation* $x_0 = 2$*, find* x_4.

(b) *What is the exact value of the root of* f*? How does this compare to our approximation* x_4 *in part (a)?*

(c) *What would happen if we chose* $x_0 = 0$ *as our initial approximation?*

Exercise 5.4.15 *Consider* $f(x) = \sin x$*. What happens when we choose* $x_0 = \pi/2$*? Explain.*

5.5 L'Hôpital's Rule

The following application of derivatives allows us to compute certain limits.

Definition 5.33: Limits of the Indeterminate Forms $\frac{0}{0}$ and $\frac{\infty}{\infty}$

A limit of a quotient $\lim\limits_{x\to a} \frac{f(x)}{g(x)}$ *is said to be an* ***indeterminate form of the type*** $\frac{0}{0}$ *if both* $f(x) \to 0$ *and* $g(x) \to 0$ *as* $x \to a$*. Likewise, it is said to be an* ***indeterminate form of the type*** $\frac{\infty}{\infty}$ *if both* $f(x) \to \pm\infty$ *and* $g(x) \to \pm\infty$ *as* $x \to a$ *(Here, the two* $\pm$ *signs are independent of each other).*

Theorem 5.34: L'Hôpital's Rule

For a limit $\lim\limits_{x\to a} \frac{f(x)}{g(x)}$ *of the indeterminate form* $\frac{0}{0}$ *or* $\frac{\infty}{\infty}$*,* $\lim\limits_{x\to a} \frac{f(x)}{g(x)} = \lim\limits_{x\to a} \frac{f'(x)}{g'(x)}$ *if* $\lim\limits_{x\to a} \frac{f'(x)}{g'(x)}$ *exists or equals* ∞ *or* $-\infty$.

This theorem is somewhat difficult to prove, in part because it incorporates so many different possibilities, so we will not prove it here.

We should also note that there may be instances where we would need to apply L'Hôpital's Rule multiple times, but we must confirm that $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ is still indeterminate before we attempt to apply L'Hôpital's Rule again. Finally, we want to mention that L'Hôpital's rule is also valid for one-sided limits and limits at infinity.

Example 5.35: L'Hôpital's Rule

Compute $\lim\limits_{x\to\pi} \dfrac{x^2 - \pi^2}{\sin x}$.

Solution. We use L'Hôpital's Rule: Since the numerator and denominator both approach zero,

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

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

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

♣

Example 5.36: L'Hôpital's Rule

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

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

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

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

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

So the original limit is 2 as well. ♣

Example 5.37: L'Hôpital's Rule

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

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

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

♣

L'Hôpital's rule concerns limits of a quotient that are indeterminate forms. But not all functions are given in the form of a quotient. But all the same, nothing prevents us from re-writing a given function in the form of a quotient. Indeed, some functions whose given form involve either a product $f(x)g(x)$ or a power $f(x)^{g(x)}$ carry indeterminacies such as $0\cdot\pm\infty$ and $1^{\pm\infty}$. Something small times something numerically large (positive or negative) could be anything. It depends on how small and how large each piece turns out to be. A number close to 1 raised to a numerically large (positive or negative) power could be anything. It depends on how close to 1 the base is, whether the base is larger than or smaller than 1, and how large the exponent is (and its sign). We can use suitable algebraic manipulations to relate them to indeterminate quotients. We will illustrate with two examples, first a product and then a power.

Example 5.38: L'Hôpital's Rule

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

Solution. This doesn't appear to be suitable for L'Hôpital's Rule, but it also is not "obvious". As x approaches zero, $\ln x$ goes to $-\infty$, so the product looks like:

$$(\text{something very small}) \cdot (\text{something very large and negative}).$$

This could be anything: it depends on *how small* and *how large* each piece of the function turns out to be. As defined earlier, this is a type of $\pm$"$0 \cdot \infty$", which is indeterminate. So we can in fact apply L'Hôpital's Rule after re-writing it in the form $\frac{\infty}{\infty}$:

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

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

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

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

Finally, we illustrate how a limit of the type "1^∞" can be indeterminate.

Example 5.39: L'Hôpital's Rule

Evaluate $\lim_{x \to 1^+} x^{1/(x-1)}$.

Solution. Plugging in $x = 1$ (from the right) gives a limit of the type "1^∞". To deal with this type of limit we will use logarithms. Let

$$L = \lim_{x \to 1^+} x^{1/(x-1)}.$$

Now, take the natural log of both sides:

$$\ln L = \lim_{x \to 1^+} \ln \left(x^{1/(x-1)} \right).$$

Using log properties we have:

$$\ln L = \lim_{x \to 1^+} \frac{\ln x}{x-1}.$$

The right side limit is now of the type $0/0$, therefore, we can apply L'Hôpital's Rule:

$$\ln L = \lim_{x \to 1^+} \frac{\ln x}{x-1} = \lim_{x \to 1^+} \frac{1/x}{1} = 1$$

Thus, $\ln L = 1$ and hence, our original limit (denoted by L) is: $L = e^1 = e$. That is,

$$L = \lim_{x \to 1^+} x^{1/(x-1)} = e.$$

In this case, even though our limit had a type of "1^∞", it actually had a value of e. ♣

Exercises for 5.5

Compute the following limits.

Exercise 5.5.1 $\lim\limits_{x\to 0}\dfrac{\cos x-1}{\sin x}$

Exercise 5.5.2 $\lim\limits_{x\to\infty}\dfrac{e^x}{x^3}$

Exercise 5.5.3 $\lim\limits_{x\to\infty}\dfrac{\ln x}{x}$

Exercise 5.5.4 $\lim\limits_{x\to\infty}\dfrac{\ln x}{\sqrt{x}}$

Exercise 5.5.5 $\lim\limits_{x\to 0}\dfrac{\sqrt{9+x}-3}{x}$

Exercise 5.5.6 $\lim\limits_{x\to 2}\dfrac{2-\sqrt{x+2}}{4-x^2}$

Exercise 5.5.7 $\lim\limits_{x\to 1}\dfrac{\sqrt{x}-1}{\sqrt[1/3]{x}-1}$

Exercise 5.5.8 $\lim\limits_{x\to 0}\dfrac{(1-x)^{1/4}-1}{x}$

Exercise 5.5.9 $\lim\limits_{t\to 0}\left(t+\dfrac{1}{t}\right)((4-t)^{3/2}-8)$

Exercise 5.5.10 $\lim\limits_{t\to 0^+}\left(\dfrac{1}{t}+\dfrac{1}{\sqrt{t}}\right)(\sqrt{t+1}-1)$

Exercise 5.5.11 $\lim\limits_{x\to 0}\dfrac{x^2}{\sqrt{2x+1}-1}$

Exercise 5.5.12 $\lim\limits_{u\to 1}\dfrac{(u-1)^3}{(1/u)-u^2+3/u-3}$

Exercise 5.5.13 $\lim\limits_{x\to 0}\dfrac{2+(1/x)}{3-(2/x)}$

Exercise 5.5.14 $\lim\limits_{x\to 0^+}\dfrac{1+5/\sqrt{x}}{2+1/\sqrt{x}}$

Exercise 5.5.15 $\displaystyle\lim_{x\to\pi/2} \frac{\cos x}{(\pi/2)-x}$

Exercise 5.5.16 $\displaystyle\lim_{x\to 0} \frac{e^x-1}{x}$

Exercise 5.5.17 $\displaystyle\lim_{x\to 0} \frac{x^2}{e^x-x-1}$

Exercise 5.5.18 $\displaystyle\lim_{x\to 1} \frac{\ln x}{x-1}$

Exercise 5.5.19 $\displaystyle\lim_{x\to 0} \frac{\ln(x^2+1)}{x}$

Exercise 5.5.20 $\displaystyle\lim_{x\to 1} \frac{x\ln x}{x^2-1}$

Exercise 5.5.21 $\displaystyle\lim_{x\to 0} \frac{\sin(2x)}{\ln(x+1)}$

Exercise 5.5.22 $\displaystyle\lim_{x\to 1} \frac{x^{1/4}-1}{x}$

Exercise 5.5.23 $\displaystyle\lim_{x\to 1} \frac{\sqrt{x}-1}{x-1}$

Exercise 5.5.24 $\displaystyle\lim_{x\to 0} \frac{3x^2+x+2}{x-4}$

Exercise 5.5.25 $\displaystyle\lim_{x\to 0} \frac{\sqrt{x+1}-1}{\sqrt{x+4}-2}$

Exercise 5.5.26 $\displaystyle\lim_{x\to 0} \frac{\sqrt{x+1}-1}{\sqrt{x+2}-2}$

Exercise 5.5.27 $\displaystyle\lim_{x\to 0^+} \frac{\sqrt{x+1}+1}{\sqrt{x+1}-1}$

Exercise 5.5.28 $\displaystyle\lim_{x\to 0} \frac{\sqrt{x^2+1}-1}{\sqrt{x+1}-1}$

Exercise 5.5.29 $\displaystyle\lim_{x\to 1} (x+5)\left(\frac{1}{2x}+\frac{1}{x+2}\right)$

Exercise 5.5.30 $\displaystyle\lim_{x\to 2} \frac{x^3-6x-2}{x^3+4}$

Exercise 5.5.31 *Discuss what happens if we try to use L'Hôpital's rule to find the limit* $\displaystyle\lim_{x\to\infty} \frac{x+\sin x}{x+1}$.

5.6 Curve Sketching

In this section, we discuss how we can tell what the graph of a function looks like by performing simple tests on its derivatives.

5.6.1. Intervals of Increase/Decrease, and the First Derivative Test

The method of Section 5.2.1 for deciding whether there is a local maximum or minimum at a critical value is not always convenient. We can instead use information about the derivative $f'(x)$ to decide; since we have already had to compute the derivative to find the critical values, there is often relatively little extra work involved in this method.

How can the derivative tell us whether there is a maximum, minimum, or neither at a point? Suppose that f is differentiable at and around $x = a$, and suppose further that a is a critical point of f. Then we have several possibilities:

1. There is a local maximum at $x = a$. This happens if $f'(x) > 0$ as we approach $x = a$ from the left (i.e. when x is in the vicinity of a, and $x < a$) and $f'(x) < 0$ as we move to the right of $x = a$ (i.e. when x is in the vicinity of a, and $x > a$).
2. There is a local minimum at $x = a$. This happens if $f'(x) < 0$ as we approach $x = a$ from the left (i.e. when x is in the vicinity of a, and $x < a$) and $f'(x) > 0$ as we move to the right of $x = a$ (i.e. when x is in the vicinity of a, and $x > a$).
3. There is neither a local maximum or local minimum at $x = a$. If $f'(x)$ does not change from negative to positive, or from positive to negative, as we move from the left of $x = a$ to the right of $x = a$ (that is, $f'(x)$ is positive on both sides of $x = a$, or negative on both sides of $x = a$) then there is neither a maximum nor minimum when $x = a$.

See the first graph in Figure 5.5 and the graph in Figure 5.6 for examples.

Example 5.40: Local Maximum and Minimum

Find all local maximum and minimum points for $f(x) = \sin x + \cos x$ using the first derivative test.

Solution. The derivative is $f'(x) = \cos x - \sin x$ and from Example 5.10 the critical values we need to consider are $\pi/4$ and $5\pi/4$.

We analyze the graphs of $\sin x$ and $\cos x$. Just to the left of $\pi/4$ the cosine is larger than the sine, so $f'(x)$ is positive; just to the right the cosine is smaller than the sine, so $f'(x)$ is negative. This means there is a local maximum at $\pi/4$. Just to the left of $5\pi/4$ the cosine is smaller than the sine, and to the right the cosine is larger than the sine. This means that the derivative $f'(x)$ is negative to the left and positive to the right, so f has a local minimum at $5\pi/4$. ♣

The above observations have obvious intuitive appeal as you examine the graphs in Figures 5.5 and 5.6. We can extend these ideas further and then formulate and prove a theorem: If the graph of f is increasing before (i.e., to the left of) $x = a$ and decreasing after (i.e., to the right of) $x = a$, then there is a local maximum at $x = a$. If the graph of f is decreasing before $x = a$ and increasing after $x = a$, then there is a local minimum at $x = a$. If the graph of f is consistently increasing on either side of $x = a$ or consistently decreasing on either side of $x = a$, then there is neither a local maximum nor a local minimum at $x = a$. We can prove the following theorem using the Mean Value Theorem.

Theorem 5.41: Intervals of Increase and Decrease

If $f'(x) > 0$ for every x in an interval, then f is increasing on that interval.
If $f'(x) < 0$ for every x in an interval, then f is decreasing on that interval.

Proof. We will prove the increasing case. The proof of the decreasing case is similar. Suppose that $f'(x) > 0$ on an interval I. Then f is differentiable, and hence also, continuous on I. If x_1 and x_2 are any two numbers in I and $x_1 < x_2$, then f is continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2). By the Mean Value Theorem, there is some c in (x_1, x_2) such that

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

But c must be in I, and thus, since $f'(x) > 0$ for every x in I, $f'(c) > 0$. Also, since $x_1 < x_2$, we have $x_2 - x_1 > 0$. Therefore, both the left hand side and the denominator of the right hand side are positive. It follows that the numerator of the right hand must be positive. That is, $f(x_2) - f(x_1) > 0$, or in other words, $f(x_1) < f(x_2)$. This shows that between x_1 and x_2 in I, the larger one, x_2, necessarily has the larger function value, $f(x_2)$, and the smaller one, x_1, necessarily have the smaller function value, $f(x_1)$. This means that f is increasing on I. ♣

Example 5.42: Local Minimum and Maximum

Consider the function $f(x) = x^4 - 2x^2$. Find where f is increasing and where f is decreasing. Use this information to find the local maximum and minimum points of f.

Solution. We compute $f'(x)$ and analyze its sign.

$$f'(x) = 4x^3 - 4x = 4x\left(x^2 - 1\right) = 4x(x-1)(x+1).$$

The solution of the inequality $f'(x) > 0$ is $(-1, 0) \cup (1, \infty)$. So, f is increasing on the interval $(-1, 0)$ and on the interval $(1, \infty)$. The solution of the inequality $f'(x) < 0$ is $(-\infty, -1) \cup (0, 1)$. So, f is decreasing on the interval $(-\infty, -1)$ and on the interval $(0, 1)$. Therefore, at the critical points -1, 0 and 1, respectively, f has a local minimum, a local maximum and a local minimum. ♣

Exercises for 5.6.1

Find all critical points and identify them as local maximum points, local minimum points, or neither.

Exercise 5.6.1 $y = x^2 - x$

Exercise 5.6.2 $y = 2 + 3x - x^3$

Exercise 5.6.3 $y = x^3 - 9x^2 + 24x$

Exercise 5.6.4 $y = x^4 - 2x^2 + 3$

Exercise 5.6.5 $y = 3x^4 - 4x^3$

Exercise 5.6.6 $y = (x^2 - 1)/x$

Exercise 5.6.7 $y = 3x^2 - (1/x^2)$

Exercise 5.6.8 $y = \cos(2x) - x$

Exercise 5.6.9 $f(x) = (5-x)/(x+2)$

Exercise 5.6.10 $f(x) = |x^2 - 121|$

Exercise 5.6.11 $f(x) = x^3/(x+1)$

Exercise 5.6.12 $f(x) = \sin^2 x$

Exercise 5.6.13 *Find the maxima and minima of* $f(x) = \sec x$.

Exercise 5.6.14 *Let* $f(\theta) = \cos^2(\theta) - 2\sin(\theta)$. *Find the intervals where* f *is increasing and the intervals where* f *is decreasing in* $[0, 2\pi]$. *Use this information to classify the critical points of* f *as either local maximums, local minimums, or neither.*

Exercise 5.6.15 *Let* $r > 0$. *Find the local maxima and minima of the function* $f(x) = \sqrt{r^2 - x^2}$ *on its domain* $[-r, r]$.

Exercise 5.6.16 *Let* $f(x) = ax^2 + bx + c$ *with* $a \neq 0$. *Show that* f *has exactly one critical point using the first derivative test. Give conditions on a and b which guarantee that the critical point will be a maximum. It is possible to see this without using calculus at all; explain.*

5.6.2. The Second Derivative Test

The basis of the first derivative test is that if the derivative changes from positive to negative at a point at which the derivative is zero then there is a local maximum at the point, and similarly for a local minimum. If f' changes from positive to negative it is decreasing; this means that the derivative of f', f'', might be negative, and if in fact f'' is negative then f' is definitely decreasing. From this we determine that there is a local maximum at the point in question. Note that f' might change from positive to negative while f'' is zero, in which case f'' gives us no information about the critical value. Similarly, if f' changes from negative to positive there is a local minimum at the point, and f' is increasing. If $f'' > 0$ at the point, this tells us that f' is increasing, and so there is a local minimum.

Example 5.43: Second Derivative

Consider again $f(x) = \sin x + \cos x$, with $f'(x) = \cos x - \sin x$ and $f''(x) = -\sin x - \cos x$. Use the second derivative test to determine which critical points are local maxima or minima.

Solution. Since $f''(\pi/4) = -\sqrt{2}/2 - \sqrt{2}/2 = -\sqrt{2} < 0$, we know there is a local maximum at $\pi/4$. Since $f''(5\pi/4) = -(-\sqrt{2}/2) - (-\sqrt{2}/2) = \sqrt{2} > 0$, there is a local minimum at $5\pi/4$. ♣

When it works, the second derivative test is often the easiest way to identify local maximum and minimum points. Sometimes the test fails, and sometimes the second derivative is quite difficult to evaluate; in such cases we must fall back on one of the previous tests.

Example 5.44: Second Derivative

Let $f(x) = x^4$ and $g(x) = -x^4$. Classify the critical points of $f(x)$ and $g(x)$ as either maximum or minimum.

Solution. The derivatives for $f(x)$ are $f'(x) = 4x^3$ and $f''(x) = 12x^2$. Zero is the only critical value, but $f''(0) = 0$, so the second derivative test tells us nothing. However, $f(x)$ is positive everywhere except at zero, so clearly $f(x)$ has a local minimum at zero.

On the other hand, for $g(x) = -x^4$, $g'(x) = -4x^3$ and $g''(x) = -12x^2$. So $g(x)$ also has zero as its only only critical value, and the second derivative is again zero, but $-x^4$ has a local maximum at zero. ♣

Exercises for 5.6.2

Find all local maximum and minimum points by the second derivative test.

Exercise 5.6.17 $y = x^2 - x$

Exercise 5.6.18 $y = 2 + 3x - x^3$

Exercise 5.6.19 $y = x^3 - 9x^2 + 24x$

Exercise 5.6.20 $y = x^4 - 2x^2 + 3$

Exercise 5.6.21 $y = 3x^4 - 4x^3$

Exercise 5.6.22 $y = (x^2 - 1)/x$

Exercise 5.6.23 $y = 3x^2 - (1/x^2)$

Exercise 5.6.24 $y = \cos(2x) - x$

Exercise 5.6.25 $y = 4x + \sqrt{1-x}$

Exercise 5.6.26 $y = (x+1)/\sqrt{5x^2+35}$

Exercise 5.6.27 $y = x^5 - x$

Exercise 5.6.28 $y = 6x + \sin 3x$

Exercise 5.6.29 $y = x + 1/x$

Exercise 5.6.30 $y = x^2 + 1/x$

Exercise 5.6.31 $y = (x+5)^{1/4}$

Exercise 5.6.32 $y = \tan^2 x$

Exercise 5.6.33 $y = \cos^2 x - \sin^2 x$

Exercise 5.6.34 $y = \sin^3 x$

5.6.3. Concavity and Inflection Points

We know that the sign of the derivative tells us whether a function is increasing or decreasing; for example, when $f'(x) > 0$, $f(x)$ is increasing. The sign of the second derivative $f''(x)$ tells us whether f' is increasing or decreasing; we have seen that if f' is zero and increasing at a point then there is a local minimum at the point. If f' is zero and decreasing at a point then there is a local maximum at the point. Thus, we extracted information about f from information about f''.

We can get information from the sign of f'' even when f' is not zero. Suppose that $f''(a) > 0$. This means that near $x = a$, f' is increasing. If $f'(a) > 0$, this means that f slopes up and is getting steeper; if $f'(a) < 0$, this means that f slopes down and is getting *less* steep. The two situations are shown in figure 5.14. A curve that is shaped like this is called **concave up**.

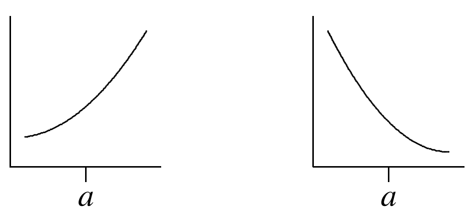


Figure 5.14: $f''(a) > 0$: $f'(a)$ positive and increasing, $f'(a)$ negative and increasing.

Now suppose that $f''(a) < 0$. This means that near $x = a$, f' is decreasing. If $f'(a) > 0$, this means that f slopes up and is getting less steep; if $f'(a) < 0$, this means that f slopes down and is getting steeper. The two situations are shown in figure 5.15. A curve that is shaped like this is called **concave down**.

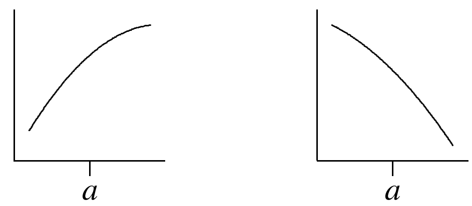


Figure 5.15: $f''(a) < 0$: $f'(a)$ positive and decreasing, $f'(a)$ negative and decreasing.

If we are trying to understand the shape of the graph of a function, knowing where it is concave up and concave down helps us to get a more accurate picture. Of particular interest are points at which the concavity changes from up to down or down to up; such points are called **inflection points**. If the concavity changes from up to down at $x = a$, f'' changes from positive to the left of a to negative to the right of a, and usually $f''(a) = 0$. We can identify such points by first finding where $f''(x)$ is zero and then checking to see whether $f''(x)$ does in fact go from positive to negative or negative to positive at these points. Note that it is possible that $f''(a) = 0$ but the concavity is the same on both sides; $f(x) = x^4$ at $x = 0$ is an example.

Example 5.45: Concavity

Describe the concavity of $f(x) = x^3 - x$.

Solution. The derivatives are $f'(x) = 3x^2 - 1$ and $f''(x) = 6x$. Since $f''(0) = 0$, there is potentially an inflection point at zero. Since $f''(x) > 0$ when $x > 0$ and $f''(x) < 0$ when $x < 0$ the concavity does change from concave down to concave up at zero, and the curve is concave down for all $x < 0$ and concave up for all $x > 0$. ♣

Note that we need to compute and analyze the second derivative to understand concavity, so we may as well try to use the second derivative test for maxima and minima. If for some reason this fails we can then try one of the other tests.

Exercises for 5.6.3

Describe the concavity of the functions below.

Exercise 5.6.35 $y = x^2 - x$

Exercise 5.6.36 $y = 2 + 3x - x^3$

Exercise 5.6.37 $y = x^3 - 9x^2 + 24x$

Exercise 5.6.38 $y = x^4 - 2x^2 + 3$

Exercise 5.6.39 $y = 3x^4 - 4x^3$

Exercise 5.6.40 $y = (x^2 - 1)/x$

Exercise 5.6.41 $y = 3x^2 - (1/x^2)$

Exercise 5.6.42 $y = \sin x + \cos x$

Exercise 5.6.43 $y = 4x + \sqrt{1-x}$

Exercise 5.6.44 $y = (x+1)/\sqrt{5x^2+35}$

Exercise 5.6.45 $y = x^5 - x$

Exercise 5.6.46 $y = 6x + \sin 3x$

Exercise 5.6.47 $y = x + 1/x$

Exercise 5.6.48 $y = x^2 + 1/x$

Exercise 5.6.49 $y = (x+5)^{1/4}$

Exercise 5.6.50 $y = \tan^2 x$

Exercise 5.6.51 $y = \cos^2 x - \sin^2 x$

Exercise 5.6.52 $y = \sin^3 x$

Exercise 5.6.53 *Identify the intervals on which the graph of the function $f(x) = x^4 - 4x^3 + 10$ is of one of these four shapes: concave up and increasing; concave up and decreasing; concave down and increasing; concave down and decreasing.*

Exercise 5.6.54 *Describe the concavity of $y = x^3 + bx^2 + cx + d$. You will need to consider different cases, depending on the values of the coefficients.*

Exercise 5.6.55 *Let n be an integer greater than or equal to two, and suppose f is a polynomial of degree n. How many inflection points can f have? Hint: Use the second derivative test and the fundamental theorem of algebra.*

5.6.4. Asymptotes and Other Things to Look For

A vertical asymptote is a place where the function becomes infinite, typically because the formula for the function has a denominator that becomes zero. For example, the reciprocal function $f(x) = 1/x$ has a vertical asymptote at $x = 0$, and the function $\tan x$ has a vertical asymptote at $x = \pi/2$ (and also at $x = -\pi/2$, $x = 3\pi/2$, etc.). Whenever the formula for a function contains a denominator it is worth looking for a vertical asymptote by checking to see if the denominator can ever be zero, and then checking the limit at such points. Note that there is not always a vertical asymptote where the derivative is zero: $f(x) = (\sin x)/x$ has a zero denominator at $x = 0$, but since $\lim_{x \to 0} (\sin x)/x = 1$ there is no asymptote there.

A horizontal asymptote is a horizontal line to which $f(x)$ gets closer and closer as x approaches ∞ (or as x approaches $-\infty$). For example, the reciprocal function has the x-axis for a horizontal asymptote. Horizontal asymptotes can be identified by computing the limits $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$. Since $\lim_{x\to\infty} 1/x = \lim_{x\to-\infty} 1/x = 0$, the line $y = 0$ (that is, the x-axis) is a horizontal asymptote in both directions.

Some functions have asymptotes that are neither horizontal nor vertical, but some other line. Such asymptotes are somewhat more difficult to identify and we will ignore them.

If the domain of the function does not extend out to infinity, we should also ask what happens as x approaches the boundary of the domain. For example, the function $y = f(x) = 1/\sqrt{r^2 - x^2}$ has domain $-r < x < r$, and y becomes infinite as x approaches either r or $-r$. In this case we might also identify this behavior because when $x = \pm r$ the denominator of the function is zero.

If there are any points where the derivative fails to exist (a cusp or corner), then we should take special note of what the function does at such a point.

Finally, it is worthwhile to notice any symmetry. A function $f(x)$ that has the same value for $-x$ as for x, i.e., $f(-x) = f(x)$, is called an "even function." Its graph is symmetric with respect to the y-axis. Some examples of even functions are: x^n when n is an even number, $\cos x$, and $\sin^2 x$. On the other hand, a function that satisfies the property $f(-x) = -f(x)$ is called an "odd function." Its graph is symmetric with respect to the origin. Some examples of odd functions are: x^n when n is an odd number, $\sin x$, and $\tan x$. Of course, most functions are neither even nor odd, and do not have any particular symmetry.

5.6.5. Summary of Curve Sketching

The following is a guideline for sketching a curve $y = f(x)$ by hand. Each item may not be relevant to the function in question, but utilizing this guideline will provide all information needed to make a detailed sketch of the function.

Guideline for Curve Sketching

1. Domain of the function
2. x- and y-Intercepts
3. Symmetry
4. Vertical and Horizontal Asymptotes
5. Intervals of Increase/Decrease, and Local Extrema
6. Concavity and Points of Inflection
7. Sketch the Graph

Example 5.46: Graph Sketching

Sketch the graph of $y = f(x)$ *where* $f(x) = \dfrac{2x^2}{x^2-1}$

Solution.

1. The domain is $\{x : x^2 - 1 \neq 0\} = \{x : x \neq \pm 1\} = (-\infty, -1) \cup (-1, 1) \cup (1, \infty)$

2. There is an x-intercept at $x = 0$. The y intercept is $y = 0$.

3. $f(-x) = f(x)$, so f is an even function (symmetric about y-axis)

4. $\displaystyle\lim_{x\to\pm\infty} \frac{2x^2}{x^2-1} = \lim_{x\to\pm\infty} \frac{2}{1-1/x^2} = 2$, so $y = 2$ is a horizontal asymptote.

   Now the denominator is 0 at $x = \pm 1$, so we compute:

   $$\lim_{x\to 1^+} \frac{2x^2}{x^2-1} = +\infty, \quad \lim_{x\to 1^-} \frac{2x^2}{x^2-1} = -\infty, \quad \lim_{x\to -1^+} \frac{2x^2}{x^2-1} = -\infty, \quad \lim_{x\to -1^-} \frac{2x^2}{x^2-1} = +\infty.$$

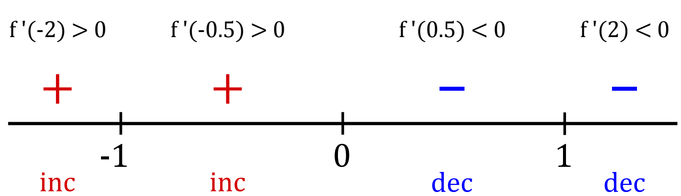
   So the lines $x = 1$ and $x = -1$ are vertical asymptotes.

5. For critical values we take the derivative:

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

   Note that $f'(x) = 0$ when $x = 0$ (the top is zero). Also, $f'(x) = DNE$ when $x = \pm 1$ (the bottom is zero). As $x = \pm 1$ is *not* in the domain of $f(x)$, the only critical number is $x = 0$ (recall that to be a critical number we need it to be in the domain of the original function).

   Drawing a number line and including *all* of the split points of $f'(x)$ we have:

   Thus f is increasing on $(-\infty, -1) \cup (-1, 0)$ and decreasing on $(0, 1) \cup (1, \infty)$.

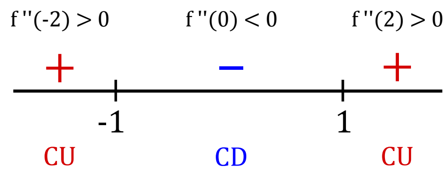
   By the first derivative test, $x = 0$ is a local max.

6. For possible inflection points we take the second derivative:

   $$f''(x) = \frac{12x^2 + 4}{(x^2-1)^3}$$

   The top is never zero. Also, the bottom is only zero when $x = \pm 1$ (neither of which are in the domain of $f(x)$). Thus, there are no possible inflection points to consider.

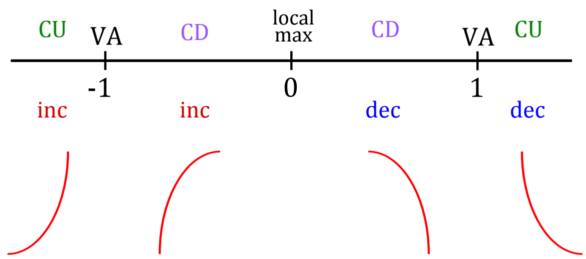
Drawing a number line and including *all* of the split points of $f''(x)$ we have:



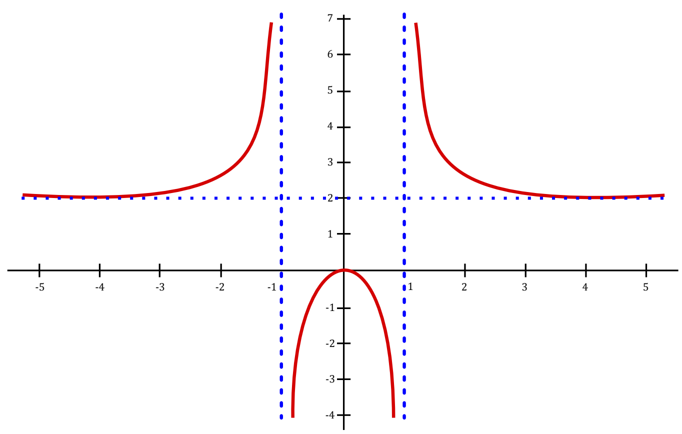
Hence f is concave up on $(-\infty,-1)\cup(1,\infty)$, concave down on $(-1,1)$.

7. We put this information together and sketch the graph.

   We combine some of this information on a single number line to see what *shape* the graph has on certain intervals:



   Note that there is a horizontal asymptote at $y=2$ and that the curve has x-int of $x=0$ and y-int of $y=0$. Therefore, a sketch of $f(x)$ is as follows:



♣

Exercises for 5.6

Sketch the curves. Identify clearly any interesting features, including local maximum and minimum points, inflection points, asymptotes, and intercepts.

Exercise 5.6.56 $y=x^5-5x^4+5x^3$

Exercise 5.6.57 $y=x^3-3x^2-9x+5$

Exercise 5.6.58 $y=(x-1)^2(x+3)^{2/3}$

Exercise 5.6.59 $x^2+x^2y^2=a^2y^2$, $a>0$.

Exercise 5.6.60 $y=xe^x$

Exercise 5.6.61 $y=(e^x+e^{-x})/2$

Exercise 5.6.62 $y=e^{-x}\cos x$

Exercise 5.6.63 $y=e^x-\sin x$

Exercise 5.6.64 $y=e^x/x$

Exercise 5.6.65 $y=4x+\sqrt{1-x}$

Exercise 5.6.66 $y=(x+1)/\sqrt{5x^2+35}$

Exercise 5.6.67 $y=x^5-x$

Exercise 5.6.68 $y=6x+\sin 3x$

Exercise 5.6.69 $y=x+1/x$

Exercise 5.6.70 $y=x^2+1/x$

Exercise 5.6.71 $y=(x+5)^{1/4}$

Exercise 5.6.72 $y=\tan^2 x$

Exercise 5.6.73 $y=\cos^2 x-\sin^2 x$

Exercise 5.6.74 $y=\sin^3 x$

Exercise 5.6.75 $y=x(x^2+1)$

Exercise 5.6.76 $y=x^3+6x^2+9x$

Exercise 5.6.77 $y=x/(x^2-9)$

Exercise 5.6.78 $y=x^2/(x^2+9)$

Exercise 5.6.79 $y=2\sqrt{x}-x$

Exercise 5.6.80 $y=3\sin(x)-\sin^3(x)$, *for* $x\in[0,2\pi]$

Exercise 5.6.81 $y=(x-1)/(x^2)$

5.7 Optimization Problems

Many important applied problems involve finding the best way to accomplish some task. Often this involves finding the maximum or minimum value of some function: the minimum time to make a certain journey, the minimum cost for doing a task, the maximum power that can be generated by a device, and so on. Many of these problems can be solved by finding the appropriate function and then using techniques of calculus to find the maximum or the minimum value required.

Generally such a problem will have the following mathematical form: Find the largest (or smallest) value of $f(x)$ when $a \leq x \leq b$. Sometimes a or b are infinite, but frequently the real world imposes some constraint on the values that x may have.

Such a problem differs in two ways from the local maximum and minimum problems we encountered when graphing functions: We are interested only in the function between a and b, and we want to know the largest or smallest value that $f(x)$ takes on, not merely values that are the largest or smallest in a small interval. That is, we seek not a local maximum or minimum but a *global* (or *absolute*) maximum or minimum.

Guidelines to solving an optimization problem.

1. Understand clearly what is to be maximized or minimized and what the contraints are.
2. Draw a diagram (if appropriate) and label it.
3. Decide what the variables are. For example, A for area, r for radius, C for cost.
4. Write a formula for the function for which you wish to find the maximum or minimum.
5. Express that formula in terms of only one variable, that is, in the form $f(x)$. Usually this is accomplished by using the given constraints.
6. Set $f'(x) = 0$ and solve. Check all critical values and endpoints to determine the extreme value(s) of $f(x)$.

Example 5.47: Largest Rectangle

Find the largest rectangle (that is, the rectangle with largest area) that fits inside the graph of the parabola $y = x^2$ below the line $y = a$ (a is an unspecified constant value), with the top side of the rectangle on the horizontal line $y = a$; see Figure 5.16.)

Solution. We want to find the maximum value of some function $A(x)$ representing area. Perhaps the hardest part of this problem is deciding what x should represent. The lower right corner of the rectangle is at (x, x^2), and once this is chosen the rectangle is completely determined. So we can let the x in $A(x)$ be the x of the parabola $f(x) = x^2$. Then the area is

$$A(x) = (2x)(a - x^2) = -2x^3 + 2ax.$$

We want the maximum value of $A(x)$ when x is in $[0,\sqrt{a}]$. (You might object to allowing $x=0$ or $x=\sqrt{a}$, since then the "rectangle" has either no width or no height, so is not "really" a rectangle. But the problem is somewhat easier if we simply allow such rectangles, which have zero area.)

Setting $0=A'(x)=6x^2+2a$ we get $x=\sqrt{a/3}$ as the only critical value. Testing this and the two endpoints, we have $A(0)=A(\sqrt{a})=0$ and $A(\sqrt{a/3})=(4/9)\sqrt{3}a^{3/2}$. The maximum area thus occurs when the rectangle has dimensions $2\sqrt{a/3}\times(2/3)a$. ♣

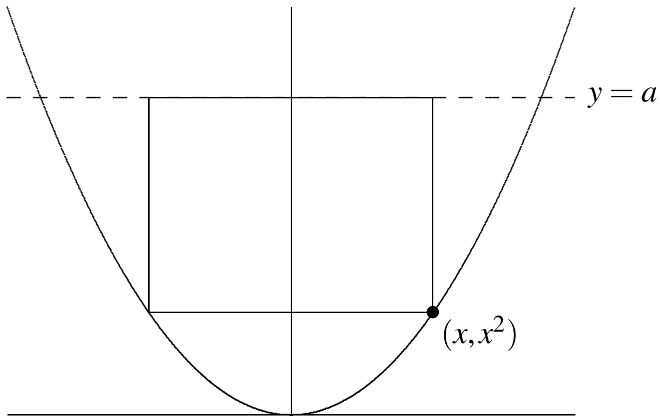


Figure 5.16: Rectangle in a parabola.

Example 5.48: Largest Cone

If you fit the largest possible cone inside a sphere, what fraction of the volume of the sphere is occupied by the cone? (Here by "cone" we mean a right circular cone, i.e., a cone for which the base is perpendicular to the axis of symmetry, and for which the cross-section cut perpendicular to the axis of symmetry at any point is a circle.)

Solution. Let R be the radius of the sphere, and let r and h be the base radius and height of the cone inside the sphere. What we want to maximize is the volume of the cone: $\pi r^2h/3$. Here R is a fixed value, but r and h can vary. Namely, we could choose r to be as large as possible—equal to R—by taking the height equal to R; or we could make the cone's height h larger at the expense of making r a little less than R. See the cross-section depicted in Figure 5.17. We have situated the picture in a convenient way relative to the x and y axes, namely, with the center of the sphere at the origin and the vertex of the cone at the far left on the x-axis.

Notice that the function we want to maximize, $\pi r^2h/3$, depends on *two* variables. This is frequently the case, but often the two variables are related in some way so that "really" there is only one variable. So our next step is to find the relationship and use it to solve for one of the variables in terms of the other, so as to have a function of only one variable to maximize. In this problem, the condition is apparent in the figure: the upper corner of the triangle, whose coordinates are $(h-R,r)$, must be on the circle of radius R. That is,

$$(h-R)^2+r^2=R^2.$$

We can solve for h in terms of r or for r in terms of h. Either involves taking a square root, but we notice that the volume function contains r^2, not r by itself, so it is easiest to solve for r^2 directly: $r^2=R^2-(h-R)^2$.

Then we substitute the result into $\pi r^2 h/3$:

$$\begin{aligned} V(h) &= \pi(R^2-(h-R)^2)h/3 \\ &= -\frac{\pi}{3}h^3+\frac{2}{3}\pi h^2 R \end{aligned}$$

We want to maximize $V(h)$ when h is between 0 and $2R$. Now we solve $0=f'(h)=-\pi h^2+(4/3)\pi hR$, getting $h=0$ or $h=4R/3$. We compute $V(0)=V(2R)=0$ and $V(4R/3)=(32/81)\pi R^3$. The maximum is the latter; since the volume of the sphere is $(4/3)\pi R^3$, the fraction of the sphere occupied by the cone is

$$\frac{(32/81)\pi R^3}{(4/3)\pi R^3}=\frac{8}{27}\approx 30\%.$$

♣

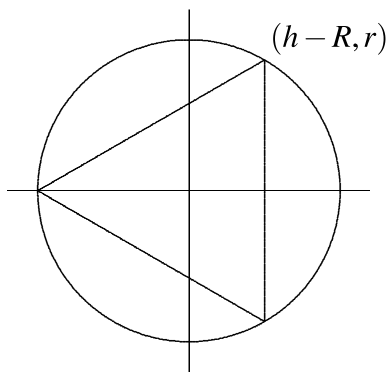


Figure 5.17: Cone in a sphere.

Example 5.49: Containers of Given Volume

You are making cylindrical containers to contain a given volume. Suppose that the top and bottom are made of a material that is N times as expensive (cost per unit area) as the material used for the lateral side of the cylinder.
Find (in terms of N) the ratio of height to base radius of the cylinder that minimizes the cost of making the containers.

Solution. Let us first choose letters to represent various things: h for the height, r for the base radius, V for the volume of the cylinder, and c for the cost per unit area of the lateral side of the cylinder; V and c are constants, h and r are variables. Now we can write the cost of materials:

$$c(2\pi rh)+Nc(2\pi r^2).$$

Again we have two variables; the relationship is provided by the fixed volume of the cylinder: $V=\pi r^2 h$. We use this relationship to eliminate h (we could eliminate r, but it's a little easier if we eliminate h, which appears in only one place in the above formula for cost). The result is

$$f(r)=2c\pi r\frac{V}{\pi r^2}+2Nc\pi r^2=\frac{2cV}{r}+2Nc\pi r^2.$$

We want to know the minimum value of this function when r is in $(0,\infty)$. We now set $0 = f'(r) = -2cV/r^2 + 4Nc\pi r$, giving $r = \sqrt[3]{V/(2N\pi)}$. Since $f''(r) = 4cV/r^3 + 4Nc\pi$ is positive when r is positive, there is a local minimum at the critical value, and hence a global minimum since there is only one critical value.

Finally, since $h = V/(\pi r^2)$,

$$\frac{h}{r} = \frac{V}{\pi r^3} = \frac{V}{\pi(V/(2N\pi))} = 2N,$$

so the minimum cost occurs when the height h is $2N$ times the radius. If, for example, there is no difference in the cost of materials, the height is twice the radius (or the height is equal to the diameter). ♣

Example 5.50: Rectangles of Given Area

Of all rectangles of area 100, which has the smallest perimeter?

Solution. First we must translate this into a purely mathematical problem in which we want to find the minimum value of a function. If x denotes one of the sides of the rectangle, then the adjacent side must be $100/x$ (in order that the area be 100). So the function we want to minimize is

$$f(x) = 2x + 2\frac{100}{x}$$

since the perimeter is twice the length plus twice the width of the rectangle. Not all values of x make sense in this problem: lengths of sides of rectangles must be positive, so $x > 0$. If $x > 0$ then so is $100/x$, so we need no second condition on x.

We next find $f'(x)$ and set it equal to zero: $0 = f'(x) = 2 - 200/x^2$. Solving $f'(x) = 0$ for x gives us $x = \pm 10$. We are interested only in $x > 0$, so only the value $x = 10$ is of interest. Since $f'(x)$ is defined everywhere on the interval $(0,\infty)$, there are no more critical values, and there are no endpoints. Is there a local maximum, minimum, or neither at $x = 10$? The second derivative is $f''(x) = 400/x^3$, and $f''(10) > 0$, so there is a local minimum. Since there is only one critical value, this is also the global minimum, so the rectangle with smallest perimeter is the 10×10 square. ♣

Example 5.51: Maximize your Profit

You want to sell a certain number n of items in order to maximize your profit. Market research tells you that if you set the price at \$1.50, you will be able to sell 5000 items, and for every 10 cents you lower the price below \$1.50 you will be able to sell another 1000 items. Suppose that your fixed costs ("start-up costs") total \$2000, and the per item cost of production ("marginal cost") is \$0.50. Find the price to set per item and the number of items sold in order to maximize profit, and also determine the maximum profit you can get.

Solution. The first step is to convert the problem into a function maximization problem. Since we want to maximize profit by setting the price per item, we should look for a function $P(x)$ representing the profit when the price per item is x. Profit is revenue minus costs, and revenue is number of items sold times the price per item, so we get $P = nx - 2000 - 0.50n$. The number of items sold is itself a function of x,

$n = 5000 + 1000(1.5 - x)/0.10$, because $(1.5 - x)/0.10$ is the number of multiples of 10 cents that the price is below \$1.50. Now we substitute for n in the profit function:

$$\begin{aligned} P(x) &= (5000 + 1000(1.5 - x)/0.10)x - 2000 - 0.5(5000 + 1000(1.5 - x)/0.10) \\ &= -10000x^2 + 25000x - 12000 \end{aligned}$$

We want to know the maximum value of this function when x is between 0 and 1.5. The derivative is $P'(x) = -20000x + 25000$, which is zero when $x = 1.25$. Since $P''(x) = -20000 < 0$, there must be a local maximum at $x = 1.25$, and since this is the only critical value it must be a global maximum as well. (Alternately, we could compute $P(0) = -12000$, $P(1.25) = 3625$, and $P(1.5) = 3000$ and note that $P(1.25)$ is the maximum of these.) Thus the maximum profit is \$3625, attained when we set the price at \$1.25 and sell 7500 items. ♣

Example 5.52: Minimize Travel Time

Suppose you want to reach a point A that is located across the sand from a nearby road (see Figure 5.18). Suppose that the road is straight, and b is the distance from A to the closest point C on the road. Let v be your speed on the road, and let w, which is less than v, be your speed on the sand. Right now you are at the point D, which is a distance a from C. At what point B should you turn off the road and head across the sand in order to minimize your travel time to A?

Solution. Let x be the distance short of C where you turn off, i.e., the distance from B to C. We want to minimize the total travel time. Recall that when traveling at constant velocity, time is distance divided by velocity.

You travel the distance $\overline{DB}$ at speed v, and then the distance $\overline{BA}$ at speed w. Since $\overline{DB} = a - x$ and, by the Pythagorean theorem, $\overline{BA} = \sqrt{x^2 + b^2}$, the total time for the trip is

$$f(x) = \frac{a - x}{v} + \frac{\sqrt{x^2 + b^2}}{w}.$$

We want to find the minimum value of f when x is between 0 and a. As usual we set $f'(x) = 0$ and solve for x:

$$\begin{aligned} 0 = f'(x) = -\frac{1}{v} + \frac{x}{w\sqrt{x^2 + b^2}} \\ w\sqrt{x^2 + b^2} &= vx \\ w^2(x^2 + b^2) &= v^2x^2 \\ w^2b^2 &= (v^2 - w^2)x^2 \\ x &= \frac{wb}{\sqrt{v^2 - w^2}} \end{aligned}$$

Notice that a does not appear in the last expression, but a is not irrelevant, since we are interested only in critical values that are in $[0, a]$, and $wb/\sqrt{v^2 - w^2}$ is either in this interval or not. If it is, we can use the second derivative to test it:

$$f''(x) = \frac{b^2}{(x^2 + b^2)^{3/2}w}.$$

Since this is always positive there is a local minimum at the critical point, and so it is a global minimum as well.

If the critical value is not in $[0, a]$ it is larger than a. In this case the minimum must occur at one of the endpoints. We can compute

$$\begin{aligned} f(0) &= \frac{a}{v} + \frac{b}{w} \\ f(a) &= \frac{\sqrt{a^2+b^2}}{w} \end{aligned}$$

but it is difficult to determine which of these is smaller by direct comparison. If, as is likely in practice, we know the values of v, w, a, and b, then it is easy to determine this. With a little cleverness, however, we can determine the minimum in general. We have seen that $f''(x)$ is always positive, so the derivative $f'(x)$ is always increasing. We know that at $wb/\sqrt{v^2-w^2}$ the derivative is zero, so for values of x less than that critical value, the derivative is negative. This means that $f(0) > f(a)$, so the minimum occurs when $x = a$.

So the upshot is this: If you start farther away from C than $wb/\sqrt{v^2-w^2}$ then you always want to cut across the sand when you are a distance $wb/\sqrt{v^2-w^2}$ from point C. If you start closer than this to C, you should cut directly across the sand. ♣

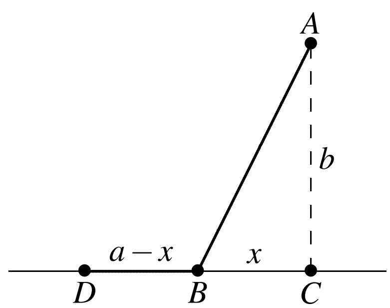


Figure 5.18: Minimizing travel time.

Exercises for Section 5.7

Exercise 5.7.1 *Find the dimensions of the rectangle of largest area having fixed perimeter* 100.

Exercise 5.7.2 *Find the dimensions of the rectangle of largest area having fixed perimeter P.*

Exercise 5.7.3 *A box with square base and no top is to hold a volume* 100. *Find the dimensions of the box that requires the least material for the five sides. Also find the ratio of height to side of the base.*

Exercise 5.7.4 *A box with square base is to hold a volume* 200. *The bottom and top are formed by folding in flaps from all four sides, so that the bottom and top consist of two layers of cardboard. Find the dimensions of the box that requires the least material. Also find the ratio of height to side of the base.*

Exercise 5.7.5 *A box with square base and no top is to hold a volume V. Find (in terms of V) the dimensions of the box that requires the least material for the five sides. Also find the ratio of height to side of the base. (This ratio will not involve V.)*

Exercise 5.7.6 *You have 100 feet of fence to make a rectangular play area alongside the wall of your house. The wall of the house bounds one side. What is the largest size possible (in square feet) for the play area?*

Exercise 5.7.7 *You have l feet of fence to make a rectangular play area alongside the wall of your house. The wall of the house bounds one side. What is the largest size possible (in square feet) for the play area?*

Exercise 5.7.8 *Marketing tells you that if you set the price of an item at \$10 then you will be unable to sell it, but that you can sell 500 items for each dollar below \$10 that you set the price. Suppose your fixed costs total \$3000, and your marginal cost is \$2 per item. What is the most profit you can make?*

Exercise 5.7.9 *Find the area of the largest rectangle that fits inside a semicircle of radius 10 (one side of the rectangle is along the diameter of the semicircle).*

Exercise 5.7.10 *Find the area of the largest rectangle that fits inside a semicircle of radius r (one side of the rectangle is along the diameter of the semicircle).*

Exercise 5.7.11 *For a cylinder with surface area 50, including the top and the bottom, find the ratio of height to base radius that maximizes the volume.*

Exercise 5.7.12 *For a cylinder with given surface area S, including the top and the bottom, find the ratio of height to base radius that maximizes the volume.*

Exercise 5.7.13 *You want to make cylindrical containers to hold 1 liter using the least amount of construction material. The side is made from a rectangular piece of material, and this can be done with no material wasted. However, the top and bottom are cut from squares of side $2r$, so that $2(2r)^2 = 8r^2$ of material is needed (rather than $2\pi r^2$, which is the total area of the top and bottom). Find the dimensions of the container using the least amount of material, and also find the ratio of height to radius for this container.*

Exercise 5.7.14 *You want to make cylindrical containers of a given volume V using the least amount of construction material. The side is made from a rectangular piece of material, and this can be done with no material wasted. However, the top and bottom are cut from squares of side $2r$, so that $2(2r)^2 = 8r^2$ of material is needed (rather than $2\pi r^2$, which is the total area of the top and bottom). Find the optimal ratio of height to radius.*

Exercise 5.7.15 *Given a right circular cone, you put an upside-down cone inside it so that its vertex is at the center of the base of the larger cone and its base is parallel to the base of the larger cone. If you choose the upside-down cone to have the largest possible volume, what fraction of the volume of the larger cone does it occupy? (Let H and R be the height and base radius of the larger cone, and let h and r be the height and base radius of the smaller cone. Hint: Use similar triangles to get an equation relating h and r.)*

Exercise 5.7.16 *A container holding a fixed volume is being made in the shape of a cylinder with a hemispherical top. (The hemispherical top has the same radius as the cylinder.) Find the ratio of height to radius of the cylinder which minimizes the cost of the container if (a) the cost per unit area of the top is twice as great as the cost per unit area of the side, and the container is made with no bottom; (b) the same as in (a), except that the container is made with a circular bottom, for which the cost per unit area is 1.5 times the cost per unit area of the side.*

Exercise 5.7.17 *A piece of cardboard is 1 meter by* $1/2$ *meter. A square is to be cut from each corner and the sides folded up to make an open-top box. What are the dimensions of the box with maximum possible volume?*

Exercise 5.7.18 *(a) A square piece of cardboard of side a is used to make an open-top box by cutting out a small square from each corner and bending up the sides. How large a square should be cut from each corner in order that the box have maximum volume? (b) What if the piece of cardboard used to make the box is a rectangle of sides a and b?*

Exercise 5.7.19 *A window consists of a rectangular piece of clear glass with a semicircular piece of colored glass on top; the colored glass transmits only* $1/2$ *as much light per unit area as the the clear glass. If the distance from top to bottom (across both the rectangle and the semicircle) is 2 meters and the window may be no more than 1.5 meters wide, find the dimensions of the rectangular portion of the window that lets through the most light.*

Exercise 5.7.20 *A window consists of a rectangular piece of clear glass with a semicircular piece of colored glass on top. Suppose that the colored glass transmits only k times as much light per unit area as the clear glass (k is between* 0 *and* 1*). If the distance from top to bottom (across both the rectangle and the semicircle) is a fixed distance H, find (in terms of k) the ratio of vertical side to horizontal side of the rectangle for which the window lets through the most light.*

Exercise 5.7.21 *You are designing a poster to contain a fixed amount A of printing (measured in square centimeters) and have margins of a centimeters at the top and bottom and b centimeters at the sides. Find the ratio of vertical dimension to horizontal dimension of the printed area on the poster if you want to minimize the amount of posterboard needed.*

Exercise 5.7.22 *What fraction of the volume of a sphere is taken up by the largest cylinder that can be fit inside the sphere?*

Exercise 5.7.23 *The U.S. post office will accept a box for shipment only if the sum of the length and girth (distance around) is at most 108 in. Find the dimensions of the largest acceptable box with square front and back.*

Exercise 5.7.24 *Find the dimensions of the lightest cylindrical can containing 0.25 liter (=250 cm*^{3}*) if the top and bottom are made of a material that is twice as heavy (per unit area) as the material used for the side.*

Exercise 5.7.25 *A conical paper cup is to hold* $1/4$ *of a liter. Find the height and radius of the cone which minimizes the amount of paper needed to make the cup. Use the formula* $\pi r\sqrt{r^2+h^2}$ *for the area of the side of a cone.*

Exercise 5.7.26 *A conical paper cup is to hold a fixed volume of water. Find the ratio of height to base radius of the cone which minimizes the amount of paper needed to make the cup. Use the formula* $\pi r\sqrt{r^2+h^2}$ *for the area of the side of a cone, called the* **lateral area** *of the cone.*

Exercise 5.7.27 *Find the fraction of the area of a triangle that is occupied by the largest rectangle that can be drawn in the triangle (with one of its sides along a side of the triangle). Show that this fraction does not depend on the dimensions of the given triangle.*

Exercise 5.7.28 *How are your answers to Problem 5.7.8 affected if the cost per item for the x items, instead of being simply \$2, decreases below \$2 in proportion to x (because of economy of scale and volume discounts) by 1 cent for each 25 items produced?*

6. Integration

6.1 Displacement and Area

Example 6.1: Object Moving in a Straight Line

An object moves in a straight line so that its speed at time t is given by $v(t) = 3t$ in, say, cm/sec. If the object is at position 10 on the straight line when $t = 0$, where is the object at any time t?

Solution. There are two reasonable ways to approach this problem. If $s(t)$ is the position of the object at time t, we know that $s'(t) = v(t)$. Based on our knowledge of derivatives, we therefore know that $s(t) = 3t^2/2 + k$, and because $s(0) = 10$ we easily discover that $k = 10$, so $s(t) = 3t^2/2 + 10$. For example, at $t = 1$ the object is at position $3/2 + 10 = 11.5$. This is certainly the easiest way to deal with this problem. Not all similar problems are so easy, as we will see; the second approach to the problem is more difficult but also more general.

We start by considering how we might approximate a solution. We know that at $t = 0$ the object is at position 10. How might we approximate its position at, say, $t = 1$? We know that the speed of the object at time $t = 0$ is 0; if its speed were constant then in the first second the object would not move and its position would still be 10 when $t = 1$. In fact, the object will not be too far from 10 at $t = 1$, but certainly we can do better. Let's look at the times 0.1, 0.2, 0.3, ..., 1.0, and try approximating the location of the object at each, by supposing that during each tenth of a second the object is going at a constant speed. Since the object initially has speed 0, we again suppose it maintains this speed, but only for a tenth of second; during that time the object would not move. During the tenth of a second from $t = 0.1$ to $t = 0.2$, we suppose that the object is traveling at 0.3 cm/sec, namely, its actual speed at $t = 0.1$. In this case the object would travel $(0.3)(0.1) = 0.03$ centimeters: 0.3 cm/sec times 0.1 seconds. Similarly, between $t = 0.2$ and $t = 0.3$ the object would travel $(0.6)(0.1) = 0.06$ centimeters. Continuing, we get as an approximation that the object travels

$$(0.0)(0.1) + (0.3)(0.1) + (0.6)(0.1) + \cdots + (2.7)(0.1) = 1.35$$

centimeters, ending up at position 11.35. This is a better approximation than 10, certainly, but is still just an approximation. (We know in fact that the object ends up at position 11.5, because we've already done the problem using the first approach.) Presumably, we will get a better approximation if we divide the time into one hundred intervals of a hundredth of a second each, and repeat the process:

$$(0.0)(0.01) + (0.03)(0.01) + (0.06)(0.01) + \cdots + (2.97)(0.01) = 1.485.$$

We thus approximate the position as 11.485. Since we know the exact answer, we can see that this is much closer, but if we did not already know the answer, we wouldn't really know how close.

We can keep this up, but we'll never really know the exact answer if we simply compute more and more examples. Let's instead look at a "typical" approximation. Suppose we divide the time into n equal intervals, and imagine that on each of these the object travels at a constant speed. Over the first time interval we approximate the distance traveled as $(0.0)(1/n)=0$, as before. During the second time interval, from $t=1/n$ to $t=2/n$, the object travels approximately $3(1/n)(1/n)=3/n^2$ centimeters. During time interval number i, the object travels approximately $(3(i-1)/n)(1/n)=3(i-1)/n^2$ centimeters, that is, its speed at time $(i-1)/n$, $3(i-1)/n$, times the length of time interval number i, $1/n$. Adding these up as before, we approximate the distance traveled as

$$(0)\frac{1}{n}+3\frac{1}{n^2}+3(2)\frac{1}{n^2}+3(3)\frac{1}{n^2}+\cdots+3(n-1)\frac{1}{n^2}$$

centimeters. What can we say about this? At first it looks rather less useful than the concrete calculations we've already done, but in fact a bit of algebra reveals it to be much more useful. We can factor out a 3 and $1/n^2$ to get

$$\frac{3}{n^2}(0+1+2+3+\cdots+(n-1)),$$

that is, $3/n^2$ times the sum of the first $n-1$ positive integers. Now we make use of a fact you may have run across before, Gauss's Equation:

$$1+2+3+\cdots+k=\frac{k(k+1)}{2}.$$

In our case we're interested in $k=n-1$, so

$$1+2+3+\cdots+(n-1)=\frac{(n-1)(n)}{2}=\frac{n^2-n}{2}.$$

This simplifies the approximate distance traveled to

$$\frac{3}{n^2}\frac{n^2-n}{2}=\frac{3}{2}\frac{n^2-n}{n^2}=\frac{3}{2}\left(\frac{n^2}{n^2}-\frac{n}{n^2}\right)=\frac{3}{2}\left(1-\frac{1}{n}\right).$$

Now this is quite easy to understand: as n gets larger and larger this approximation gets closer and closer to $(3/2)(1-0)=3/2$, so that $3/2$ is the exact distance traveled during one second, and the final position is 11.5.

So for $t=1$, at least, this rather cumbersome approach gives the same answer as the first approach. But really there's nothing special about $t=1$; let's just call it t instead. In this case the approximate distance traveled during time interval number i is $3(i-1)(t/n)(t/n)=3(i-1)t^2/n^2$, that is, speed $3(i-1)(t/n)$ times time t/n, and the total distance traveled is approximately

$$(0)\frac{t}{n}+3(1)\frac{t^2}{n^2}+3(2)\frac{t^2}{n^2}+3(3)\frac{t^2}{n^2}+\cdots+3(n-1)\frac{t^2}{n^2}.$$

As before we can simplify this to

$$\frac{3t^2}{n^2}(0+1+2+\cdots+(n-1))=\frac{3t^2}{n^2}\frac{n^2-n}{2}=\frac{3}{2}t^2\left(1-\frac{1}{n}\right).$$

In the limit, as n gets larger, this gets closer and closer to $(3/2)t^2$ and the approximated position of the object gets closer and closer to $(3/2)t^2+10$, so the actual position is $(3/2)t^2+10$, exactly the answer given by the first approach to the problem. ♣

Example 6.2: Area under the Line

Find the area under the curve $y = 3x$ between $x = 0$ and any positive value x.

Solution. There is here no obvious analogue to the first approach in the previous example, but the second approach works fine. (Since the function $y = 3x$ is so simple, there is another approach that works here, but it is even more limited in potential application than is approach number one.) How might we approximate the desired area? We know how to compute areas of rectangles, so we approximate the area by rectangles. Jumping straight to the general case, suppose we divide the interval between 0 and x into n equal subintervals, and use a rectangle above each subinterval to approximate the area under the curve. There are many ways we might do this, but let's use the height of the curve at the left endpoint of the subinterval as the height of the rectangle, as in figure 6.1. The height of rectangle number i is then $3(i-1)(x/n)$, the width is x/n, and the area is $3(i-1)(x^2/n^2)$. The total area of the rectangles is

$$(0)\frac{x}{n} + 3(1)\frac{x^2}{n^2} + 3(2)\frac{x^2}{n^2} + 3(3)\frac{x^2}{n^2} + \cdots + 3(n-1)\frac{x^2}{n^2}.$$

By factoring out $3x^2/n^2$ this simplifies to

$$\frac{3x^2}{n^2}(0+1+2+\cdots+(n-1)) = \frac{3x^2}{n^2}\frac{n^2-n}{2} = \frac{3}{2}x^2\left(1-\frac{1}{n}\right).$$

As n gets larger this gets closer and closer to $3x^2/2$, which must therefore be the true area under the curve. ♣

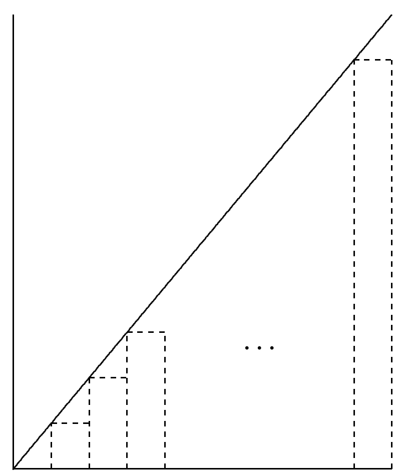


Figure 6.1: Approximating the area under $y = 3x$ with rectangles.

What you will have noticed, of course, is that while the problem in the second example appears to be much different than the problem in the first example, and while the easy approach to problem one does not appear to apply to problem two, the "approximation" approach works in both, and moreover the *calculations are identical.* As we will see, there are many, many problems that appear much different on the surface but turn out to be the same as these problems, in the sense that when we try to approximate solutions we end up with mathematics that looks like the two examples, though of course the function involved will not always be so simple.

Even better, we now see that while the second problem did not appear to be amenable to approach one, it can in fact be solved in the same way. The reasoning is this: we know that problem one can be solved easily by finding a function whose derivative is $3t$. We also know that mathematically the two problems are the same, because both can be solved by taking a limit of a sum, and the sums are identical. Therefore, we don't really need to compute the limit of either sum because we know that we will get the same answer by computing a function with the derivative $3t$ or, which is the same thing, $3x$.

It's true that the first problem had the added complication of the "10", and we certainly need to be able to deal with such minor variations, but that turns out to be quite simple. The lesson then is this: whenever we can solve a problem by taking the limit of a sum of a certain form, instead of computing the (often nasty) limit we can find a new function with a certain derivative.

6.1.1. Riemann Sums

A fundamental calculus technique is to first answer a given problem with an approximation, then refine that approximation to make it better, then use limits in the refining process to find the exact answer. That is exactly what we will do here to develop a technique to find the area of more complicated regions.

Consider the region given in Figure 6.2, which is the area under $y = 4x - x^2$ on $[0,4]$. What is the signed area of this region? While we will not use this notation in this section, we will soon see that this is equivalent to finding the integral given by $\int_0^4 (4x - x^2)\, dx$,

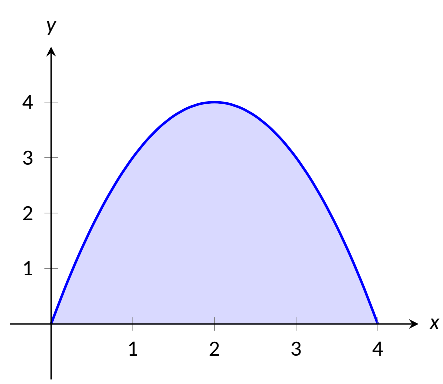


Figure 6.2: $f(x) = 4x - x^2$

We start by approximating. We can surround the region with a rectangle with height and width of

4 and find the area is approximately 16 square units. This is obviously an *over–approximation*; we are including area in the rectangle that is not under the parabola. How can we refine our approximation to make it better? The key to this section is this answer: *use more rectangles.*

Let's use four rectangles of equal width of 1. This partitions the interval $[0,4]$ into four *subintervals*, $[0,1]$, $[1,2]$, $[2,3]$ and $[3,4]$. On each subinterval we will draw a rectangle.

There are three common ways to determine the height of these rectangles: the **Left Hand Rule**, the **Right Hand Rule**, and the **Midpoint Rule**. The **Left Hand Rule** says to evaluate the function at the left–hand endpoint of the subinterval and make the rectangle that height. In Figure 6.3 below, the rectangle drawn on the interval $[2,3]$ has height determined by the Left Hand Rule; it has a height of $f(2) = 4$. (The rectangle is labeled "LHR.")

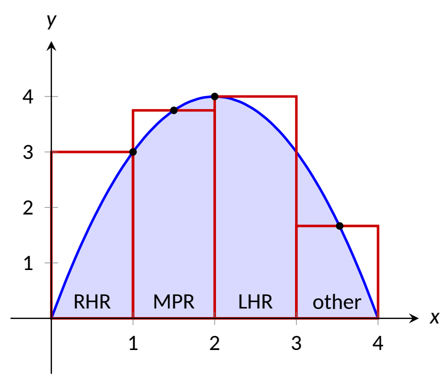


Figure 6.3: Approximating area using rectangles

The **Right Hand Rule** says the opposite: on each subinterval, evaluate the function at the right endpoint and make the rectangle that height. In the figure, the rectangle drawn on $[0,1]$ is drawn using $f(1) = 3$ as its height; this rectangle is labeled "RHR.".

The **Midpoint Rule** says that on each subinterval, evaluate the function at the midpoint and make the rectangle that height. The rectangle drawn on $[1,2]$ was made using the Midpoint Rule, with a height of $f(1.5) = 3.75$. That rectangle is labeled "MPR."

These are the three most common rules for determining the heights of approximating rectangles, but we are not forced to use one of these three methods. The rectangle on $[3,4]$ has a height of approximately $f(3.53)$, very close to the Midpoint Rule. It was chosen so that the area of the rectangle is *exactly* the area of the region under f on $[3,4]$. (Later you'll be able to figure how to do this, too.)

The following example will approximate the area under $f(x) = 4x - x^2$ using these rules.

Example 6.3: Using the Left Hand, Right Hand and Midpoint Rules

Approximate the area under $f(x) = 4x - x^2$ on the interval $[0,4]$ using the Left Hand Rule, the Right Hand Rule, and the Midpoint Rule, using four equally spaced subintervals.

Solution. We break the interval $[0,4]$ into four subintervals as before. In Figure 6.4 we see four rectangles drawn on $f(x) = 4x - x^2$ using the Left Hand Rule. (The areas of the rectangles are given in each figure.)

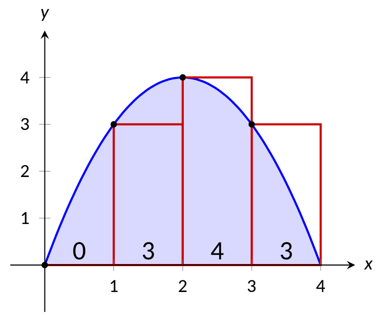


Figure 6.4: Approximating area using the Left Hand Rule

Note how in the first subinterval, $[0,1]$, the rectangle has height $f(0)=0$. We add up the areas of each rectangle (height$\times$ width) for our Left Hand Rule approximation:

$$\begin{aligned} f(0)\cdot 1+f(1)\cdot 1+f(2)\cdot 1+f(3)\cdot 1 &= \\ 0+3+4+3 &= 10. \end{aligned}$$

Figure 6.5 shows four rectangles drawn under $f(x)$ using the Right Hand Rule; note how the $[3,4]$ subinterval has a rectangle of height 0.

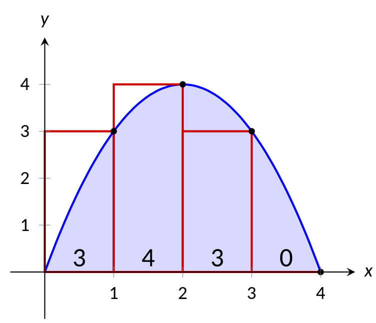


Figure 6.5: Approximating area using the Right Hand Rule,

In this figure, these rectangles seem to be the mirror image of those found in Figure 6.4. (This is because of the symmetry of our shaded region.) Our approximation gives the same answer as before, though calculated a different way:

$$\begin{aligned} f(1)\cdot 1+f(2)\cdot 1+f(3)\cdot 1+f(4)\cdot 1 &= \\ 3+4+3+0 &= 10. \end{aligned}$$

Figure 6.6 shows four rectangles drawn under $f(x)$ using the Midpoint Rule.

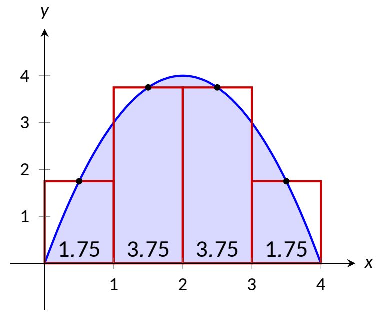


Figure 6.6: Approximating area using the Midpoint Rule

This gives an approximation of the area as:

$$f(0.5)\cdot 1 + f(1.5)\cdot 1 + f(2.5)\cdot 1 + f(3.5)\cdot 1 =$$
$$1.75 + 3.75 + 3.75 + 1.75 = 11.$$

Our three methods provide two approximations of the area under $f(x) = 4x - x^2$: 10 and 11. ♣

It is hard to tell at this moment which is a better approximation: 10 or 11? We can continue to refine our approximation by using more rectangles. The notation can become unwieldy, though, as we add up longer and longer lists of numbers. We introduce **summation notation** (also called **sigma notation**) to solve this problem.

Suppose we wish to add up a list of numbers $a_1, a_2, a_3, \ldots, a_9$. Instead of writing

$$a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9,$$

we use summation notation and write

$$\sum_{i=1}^{9} a_i$$

The upper case sigma ($\sum$) represents the term "sum" and a_i is referred to as the summand. The index of summation in this example is i; any symbol can be used. By convention, the index takes on only the integer values between (and including) the lower and upper bounds, here equal to 1 and 9 respectively.

Let's practice using this notation.

Example 6.4: Using Summation Notation

Let the numbers $\{a_i\}$ be defined as $a_i = 2i - 1$ for integers i, where $i \geq 1$. So $a_1 = 1$, $a_2 = 3$, $a_3 = 5$, etc. (The output is the positive odd integers). Evaluate the following summations:

1. $\sum_{i=1}^{6} a_i$
2. $\sum_{i=3}^{7} (3a_i - 4)$
3. $\sum_{i=1}^{4} (a_i)^2$

Solution.

1.

$$\begin{aligned}\sum_{i=1}^{6} a_i &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6 \\ &= 1+3+5+7+9+11 \\ &= 36.\end{aligned}$$

2. Note the starting value is different than 1:

$$\begin{aligned}\sum_{i=3}^{7}(3a_i - 4) &= (3a_3-4)+(3a_4-4)+(3a_5-4)+(3a_6-4)+(3a_7-4) \\ &= 11+17+23+29+35 \\ &= 115.\end{aligned}$$

3.

$$\begin{aligned}\sum_{i=1}^{4}(a_i)^2 &= (a_1)^2+(a_2)^2+(a_3)^2+(a_4)^2 \\ &= 1^2+3^2+5^2+7^2 \\ &= 84\end{aligned}$$

♣

The following theorem gives some properties of summations that allow us to work with them without writing individual terms. Examples will follow.

Theorem 6.5: Properties of Summations

1. $\displaystyle\sum_{i=1}^{n} c = c\cdot n$, *where c is a constant.*
2. $\displaystyle\sum_{i=m}^{n}(a_i \pm b_i) = \sum_{i=m}^{n} a_i \pm \sum_{i=m}^{n} b_i$
3. $\displaystyle\sum_{i=m}^{n} c\cdot a_i = c\cdot \sum_{i=m}^{n} a_i$
4. $\displaystyle\sum_{i=m}^{j} a_i + \sum_{i=j+1}^{n} a_i = \sum_{i=m}^{n} a_i$
5. $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$
6. $\displaystyle\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
7. $\displaystyle\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$

Example 6.6: Evaluating Summations

Use Theorem 6.5 to evaluate

$$\sum_{i=1}^{6} a_i = \sum_{i=1}^{6}(2i-1)$$

where i are integers and $i \geq 1$.

Solution.

$$\begin{aligned}\sum_{i=1}^{6}(2i-1) &= \sum_{i=1}^{6}2i - \sum_{i=1}^{6}(1)\\ &= \left(2\sum_{i=1}^{6}i\right) - 6\\ &= 2\frac{6(6+1)}{2} - 6\\ &= 42-6=36\end{aligned}$$

We obtained the same answer without writing out all six terms. When dealing with small values of n, it may be faster to write the terms out by hand. However, Theorem 6.5 is incredibly important when dealing with large sums as we'll soon see. ♣

Consider again $f(x) = 4x - x^2$. We will approximate the area under this curve (again for $[0,4]$ using sixteen equally spaced subintervals and the Right Hand Rule. Before doing so, we will do some careful preparation.

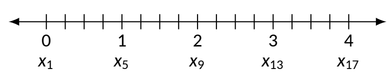


Figure 6.7: Dividing $[0,4]$ into 16 equally spaced subintervals

Figure 6.7 shows a number line of $[0,4]$ divided into sixteen equally spaced subintervals. We denote 0 as x_1; we have marked the values of x_5, x_9, x_{13} and x_{17}. We could mark them all, but the figure would get crowded. While it is easy to figure that $x_{10} = 2.25$, in general, we want a method of determining the value of x_i without consulting the figure. Consider:

$$\begin{array}{rcl} x_i &=& x_1 + (i-1)\Delta x \\ \textit{where} && \\ x_1 &:& \text{starting value} \\ (i-1) &:& \text{number of subintervals between } x_1 \text{ and } x_i \\ \Delta x &:& \text{subinterval width} \end{array}$$

So $x_{10} = x_1 + 9(4/16) = 2.25$.

If we had partitioned $[0,4]$ into 100 equally spaced subintervals, each subinterval would have length $\Delta x = 4/100 = 0.04$. We could compute x_{32} as

$$x_{32} = 0 + 31(4/100) = 1.24.$$

(That was far faster than creating a sketch first.)

Given *any* subdivision of $[0,4]$, the first subinterval is $[x_1, x_2]$; the second is $[x_2, x_3]$; the i^{th} subinterval is $[x_i, x_{i+1}]$.

When using the Left Hand Rule, the height of the i^{th} rectangle will be $f(x_i)$.
When using the Right Hand Rule, the height of the i^{th} rectangle will be $f(x_{i+1})$.
When using the Midpoint Rule, the height of the i^{th} rectangle will be $f\left(\dfrac{x_i+x_{i+1}}{2}\right)$.

Thus approximating the area under $f(x)=4x-x^2$ on $[0,4]$ with sixteen equally spaced subintervals can be expressed as follows, where $\Delta x = 4/16 = 1/4$:

Left Hand Rule: $\displaystyle\sum_{i=1}^{16} f(x_i)\Delta x$

Right Hand Rule: $\displaystyle\sum_{i=1}^{16} f(x_{i+1})\Delta x$

Midpoint Rule: $\displaystyle\sum_{i=1}^{16} f\left(\frac{x_i+x_{i+1}}{2}\right)\Delta x$

We use these formulas in the following example.

Example 6.7: Approximating Area Using Sums

Approximate the area under $f(x)=4x-x^2$ on $[0,4]$ using the Right Hand Rule and summation formulas with sixteen and 1000 equally spaced intervals.

Solution. Using sixteen equally spaced intervals and the Right Hand Rule, we can approximate the area as

$$\sum_{i=1}^{16} f(x_{i+1})\Delta x.$$

We have $\Delta x = 4/16 = 0.25$. Since $x_i = 0+(i-1)\Delta x$, we have

$$\begin{aligned} x_{i+1} &= 0+\big((i+1)-1\big)\Delta x \\ &= i\Delta x \end{aligned}$$

Using the summation formulas, consider:

$$\begin{aligned} \sum_{i=1}^{16} f(x_{i+1})\Delta x &= \sum_{i=1}^{16} f(i\Delta x)\Delta x \\ &= \sum_{i=1}^{16} \big(4i\Delta x-(i\Delta x)^2\big)\Delta x \\ &= \sum_{i=1}^{16} (4i\Delta x^2 - i^2\Delta x^3) \\ &= (4\Delta x^2)\sum_{i=1}^{16} i - \Delta x^3 \sum_{i=1}^{16} i^2 \qquad (6.1) \\ &= (4\Delta x^2)\frac{16\cdot 17}{2} - \Delta x^3 \frac{16(17)(33)}{6} \end{aligned}$$

$$= 4 \cdot 0.25^2 \cdot 136 - 0.25^3 \cdot 1496$$
$$= 10.625$$

We were able to sum up the areas of sixteen rectangles with very little computation. In Figure 6.8 the function and the sixteen rectangles are graphed. While some rectangles over–approximate the area, other under–approximate the area (by about the same amount). Thus our approximate area of 10.625 is likely a fairly good approximation.

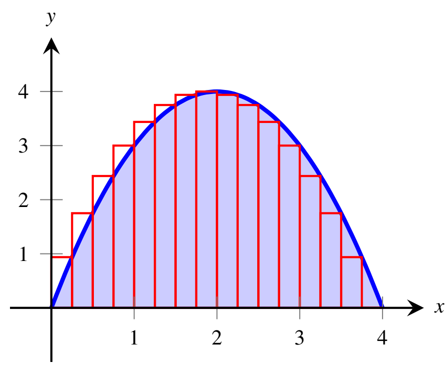


Figure 6.8: Approximating area with the Right Hand Rule and 16 evenly spaced subintervals

Notice Equation (6.1); by changing the 16's to 1,000's (and appropriately changing the value of Δx), we can use that equation to sum up 1000 rectangles!

We do so here, skipping from the original summand to the equivalent of Equation (6.1) to save space. Note that $\Delta x = 4/1000 = 0.004$.

$$\begin{aligned}\sum_{i=1}^{1000} f(x_{i+1})\Delta x &= (4\Delta x^2)\sum_{i=1}^{1000} i - \Delta x^3 \sum_{i=1}^{1000} i^2 \\ &= (4\Delta x^2)\frac{1000 \cdot 1001}{2} - \Delta x^3 \frac{1000(1001)(2001)}{6} \\ &= 4 \cdot 0.004^2 \cdot 500500 - 0.004^3 \cdot 333,833,500 \\ &= 10.666656\end{aligned}$$

Using many, many rectangles, we have a likely good approximation of the area under $f(x) = 4x - x^2$ of ≈ 10.666656. ♣

Before the above example, we stated the summations for the Left Hand, Right Hand and Midpoint Rules. Each had the same basic structure, which was:

1. each rectangle has the same width, which we referred to as Δx, and
2. each rectangle's height is determined by evaluating $f(x)$ at a particular point in each subinterval. For instance, the Left Hand Rule states that each rectangle's height is determined by evaluating $f(x)$ at the left hand endpoint of the subinterval the rectangle lives on.

One could partition an interval $[a, b]$ with subintervals that did not have the same width. We refer to the length of the first subinterval as Δx_1, the length of the second subinterval as Δx_2, and so on, giving the length of the i^{th} subinterval as Δx_i. Also, one could determine each rectangle's height by evaluating $f(x)$ at *any* point in the i^{th} subinterval. We refer to the point picked in the first subinterval as c_1, the point picked in the second subinterval as c_2, and so on, with c_i representing the point picked in the i^{th} subinterval. Thus the height of the i^{th} subinterval would be $f(c_i)$, and the area of the i^{th} rectangle would be $f(c_i)\Delta x_i$.

Summations of rectangles with area $f(c_i)\Delta x_i$ are named after mathematician Georg Friedrich Bernhard Riemann, as given in the following definition.

Definition 6.8: Riemann Sum

Let $f(x)$ be defined on the closed interval $[a,b]$ and let Δx be a partition of $[a,b]$, with

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b.$$

Let Δx_i denote the length of the i^{th} subinterval $[x_i, x_{i+1}]$ and let c_i denote any value in the i^{th} subinterval.
The sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i$$

*is a **Riemann sum** of $f(x)$ on $[a,b]$.*

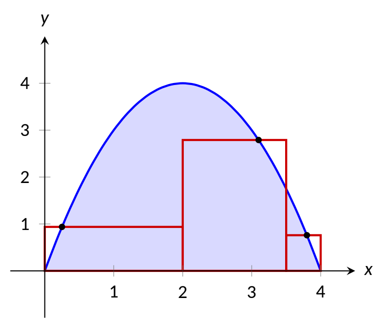


Figure 6.9: General Riemann sum to approximate the area under $f(x) = 4x - x^2$

Figure 6.9 shows the approximating rectangles of a Riemann sum. While the rectangles in this example do not approximate well the shaded area, they demonstrate that the subinterval widths may vary and the heights of the rectangles can be determined without following a particular rule.

Riemann sums are typically calculated using one of the three rules we have introduced. The uniformity of construction makes computations easier. Before working another example, let's summarize some of what we have learned in a convenient way.

Riemann Sums

Consider a function $f(x)$ defined on an interval $[a,b]$. The area under this curve is approximated by $\sum_{i=1}^{n} f(c_i)\Delta x_i$.

1. When the n subintervals have equal length, $\Delta x_i = \Delta x = \dfrac{b-a}{n}$.
2. The i^{th} term of the partition is $x_i = a+(i-1)\Delta x$. (This makes $x_{n+1} = b$.)
3. The Left Hand Rule summation is: $\displaystyle\sum_{i=1}^{n} f(x_i)\Delta x$.
4. The Right Hand Rule summation is: $\displaystyle\sum_{i=1}^{n} f(x_{i+1})\Delta x$.
5. The Midpoint Rule summation is: $\displaystyle\sum_{i=1}^{n} f\left(\frac{x_i+x_{x+1}}{2}\right)\Delta x$.

Let's do another example.

Example 6.9: Approximating Area Using Sums

Approximate the area under $f(x) = (5x+2)$ on the interval $[-2,3]$ using the Midpoint Rule and ten equally spaced intervals.

Solution. Following the above discussion, we have

$$\Delta x = \frac{3-(-2)}{10} = 1/2$$
$$x_i = (-2)+(1/2)(i-1) = i/2-5/2.$$

As we are using the Midpoint Rule, we will also need x_{i+1} and $\dfrac{x_i+x_{i+1}}{2}$. Since $x_i = i/2-5/2$, $x_{i+1} = (i+1)/2-5/2 = i/2-2$. This gives

$$\frac{x_i+x_{i+1}}{2} = \frac{(i/2-5/2)+(i/2-2)}{2} = \frac{i-9/2}{2} = i/2-9/4.$$

We now construct the Riemann sum and compute its value using summation formulas.

$$\begin{aligned}\sum_{i=1}^{10} f\left(\frac{x_i+x_{i+1}}{2}\right)\Delta x &= \sum_{i=1}^{10} f(i/2-9/4)\Delta x\\ &= \sum_{i=1}^{10}\big(5(i/2-9/4)+2\big)\Delta x\\ &= \Delta x\sum_{i=1}^{10}\left[\left(\frac{5}{2}\right)i-\frac{37}{4}\right]\end{aligned}$$

$$
\begin{aligned}
&= \Delta x\left(\frac{5}{2}\sum_{i=1}^{10}(i) - \sum_{i=1}^{10}\left(\frac{37}{4}\right)\right)\\
&= \frac{1}{2}\left(\frac{5}{2}\cdot\frac{10(11)}{2} - 10\cdot\frac{37}{4}\right)\\
&= \frac{45}{2} = 22.5
\end{aligned}
$$

Note the graph of $f(x) = 5x+2$ in Figure 6.10. The regions whose areas are computed are triangles, meaning we can find the exact answer without summation techniques. We find that the exact answer is indeed 22.5. One of the strengths of the Midpoint Rule is that often each rectangle includes area that should not be counted, but misses other area that should. When the partition width is small, these two amounts are about equal and these errors almost "cancel each other out." In this example, since our function is a line, these errors are exactly equal and they do cancel each other out, giving us the exact answer.

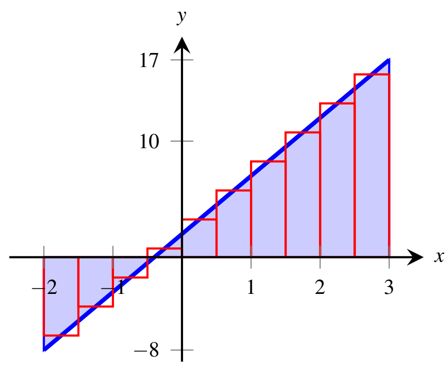


Figure 6.10: Approximating area using the Midpoint Rule and 10 evenly spaced subintervals

Note too that when the function is negative, the rectangles have a "negative" height. When we compute the area of the rectangle, we use $f(c_i)\Delta x$; when f is negative, the area is counted as negative. ♣

Notice in the previous example that while we used ten equally spaced intervals, the number "10" didn't play a big role in the calculations until the very end. Mathematicians love abstract ideas; let's approximate the area of another region using n subintervals, where we do not specify a value of n until the very end.

Example 6.10: Approximating Area Using Sums

Revisit $f(x) = 4x - x^2$ on the interval $[0,4,]$ yet again. Approximate the area under this curve using the Right Hand Rule with n equally spaced subintervals.

Solution. We know $\Delta x = \frac{4-0}{n} = 4/n$. We also find $x_i = 0 + \Delta x(i-1) = 4(i-1)/n$. The Right Hand Rule uses x_{i+1}, which is $x_{i+1} = 4i/n$. We construct the Right Hand Rule Riemann sum as follows.

$$
\begin{aligned}
\sum_{i=1}^{n} f(x_{i+1})\Delta x &= \sum_{i=1}^{n} f\left(\frac{4i}{n}\right)\Delta x\\
&= \sum_{i=1}^{n}\left[4\frac{4i}{n} - \left(\frac{4i}{n}\right)^2\right]\Delta x\\
&= \sum_{i=1}^{n}\left(\frac{16\Delta x}{n}\right)i - \sum_{i=1}^{n}\left(\frac{16\Delta x}{n^2}\right)i^2
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{16\Delta x}{n}\right)\sum_{i=1}^{n} i - \left(\frac{16\Delta x}{n^2}\right)\sum_{i=1}^{n} i^2 \\
&= \left(\frac{16\Delta x}{n}\right)\cdot\frac{n(n+1)}{2} - \left(\frac{16\Delta x}{n^2}\right)\frac{n(n+1)(2n+1)}{6} \qquad \left(\begin{array}{c}\text{recall}\\ \Delta x = 4/n\end{array}\right) \\
&= \frac{32(n+1)}{n} - \frac{32(n+1)(2n+1)}{3n^2} \qquad \text{(now simplify)} \\
&= \frac{32}{3}\left(1-\frac{1}{n^2}\right)
\end{aligned}
$$

The result is an amazing, easy to use formula. To approximate the area with ten equally spaced subintervals and the Right Hand Rule, set $n = 10$ and compute

$$\frac{32}{3}\left(1-\frac{1}{10^2}\right) = 10.56.$$

Recall how earlier we approximated the area with 4 subintervals; with $n = 4$, the formula gives 10, our answer as before.

It is now easy to approximate the area with 1,000,000 subintervals! Hand-held calculators will round off the answer a bit prematurely giving an answer of 10.66666667. (The actual answer is 10.666666666656.)

We now take an important leap. Up to this point, our mathematics has been limited to geometry and algebra (finding areas and manipulating expressions). Now we apply *calculus*. For any *finite* n, we know that the corresponding Right Hand Rule Riemann sum is:

$$\frac{32}{3}\left(1-\frac{1}{n^2}\right).$$

Both common sense and high–level mathematics tell us that as n gets large, the approximation gets better. In fact, if we take the *limit* as $n \to \infty$, we get the *exact area*. That is,

$$
\begin{aligned}
\lim_{n\to\infty}\frac{32}{3}\left(1-\frac{1}{n^2}\right) &= \frac{32}{3}(1-0) \\
&= \frac{32}{3} = 10.\overline{6}
\end{aligned}
$$

This is a fantastic result. By considering n equally–spaced subintervals, we obtained a formula for an approximation of the area that involved our variable n. As n grows large – without bound – the error shrinks to zero and we obtain the exact area. ♣

This section started with a fundamental calculus technique: make an approximation, refine the approximation to make it better, then use limits in the refining process to get an exact answer. That is precisely what we just did.

Let's practice this again.

Example 6.11: Approximating Area With a Formula, Using Sums

Find a formula that approximates the area under $f(x) = x^3$ on the interval $[-1,5]$ using the Right Hand Rule and n equally spaced subintervals, then take the limit as $n \to \infty$ to find the exact area.

Solution. We have $\Delta x = \frac{5-(-1)}{n} = 6/n$. We have $x_i = (-1) + (i-1)\Delta x$; as the Right Hand Rule uses x_{i+1}, we have $x_{i+1} = (-1) + i\Delta x$.

The Riemann sum corresponding to the Right Hand Rule is (followed by simplifications):

$$
\begin{aligned}
\sum_{i=1}^{n} f(x_{i+1})\Delta x &= \sum_{i=1}^{n} f(-1 + i\Delta x)\Delta x \\
&= \sum_{i=1}^{n} (-1 + i\Delta x)^3 \Delta x \\
&= \sum_{i=1}^{n} \big((i\Delta x)^3 - 3(i\Delta x)^2 + 3i\Delta x - 1\big)\Delta x \\
&= \sum_{i=1}^{n} \big(i^3\Delta x^4 - 3i^2\Delta x^3 + 3i\Delta x^2 - \Delta x\big) \\
&= \Delta x^4 \sum_{i=1}^{n} i^3 - 3\Delta x^3 \sum_{i=1}^{n} i^2 + 3\Delta x^2 \sum_{i=1}^{n} i - \sum_{i=1}^{n} \Delta x \\
&= \Delta x^4 \left(\frac{n(n+1)}{2}\right)^2 - 3\Delta x^3 \frac{n(n+1)(2n+1)}{6} + 3\Delta x^2 \frac{n(n+1)}{2} - n\Delta x \\
&= \frac{1296}{n^4} \cdot \frac{n^2(n+1)^2}{4} - 3\frac{216}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} + 3\frac{36}{n^2}\frac{n(n+1)}{2} - 6 \\
&= 156 + \frac{378}{n} + \frac{216}{n^2}
\end{aligned}
$$

Once again, we have found a compact formula for approximating the area with n equally spaced subintervals and the Right Hand Rule. Using ten subintervals, we have an approximation of 195.96 (these rectangles are shown in Figure 6.11). Using $n = 100$ gives an approximation of 159.802.

Now find the exact answer using a limit:

$$
\lim_{n\to\infty} \left(156 + \frac{378}{n} + \frac{216}{n^2}\right) = 156.
$$

♣

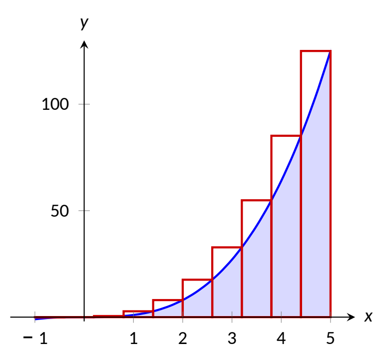


Figure 6.11: Approximating area using the Right Hand Rule and 10 evenly spaced subintervals.

We have used limits to evaluate exactly given definite limits. Will this always work? We will show, given not–very–restrictive conditions, that yes, it will always work.

The previous two examples demonstrated how an expression such as

$$\sum_{i=1}^{n} f(x_{i+1})\Delta x$$

can be rewritten as an expression explicitly involving n, such as $32/3(1-1/n^2)$.

Viewed in this manner, we can think of the summation as a function of n. An n value is given (where n is a positive integer), and the sum of areas of n equally spaced rectangles is returned, using the Left Hand, Right Hand, or Midpoint Rules.

Given a function $f(x)$ defined on the interval $[a,b]$ let:

- $S_L(n) = \sum_{i=1}^{n} f(x_i)\Delta x$, the sum of equally spaced rectangles formed using the Left Hand Rule,
- $S_R(n) = \sum_{i=1}^{n} f(x_{i+1})\Delta x$, the sum of equally spaced rectangles formed using the Right Hand Rule, and
- $S_M(n) = \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x$, the sum of equally spaced rectangles formed using the Midpoint Rule.

Recall the definition of a limit as $n \to \infty$: $\lim_{n\to\infty} S_L(n) = K$ if, given any $\varepsilon > 0$, there exists $\delta > 0$ such that $|S_L(n) - K| < \varepsilon$ whenever $n \geq \delta$.

The following theorem states that we can use any of our three rules to find the exact value of the area under $f(x)$ on $[a,b]$. It also goes two steps further. The theorem states that the height of each rectangle doesn't have to be determined following a specific rule, but could be $f(c_i)$, where c_i is any point in the i^{th} subinterval, as discussed earlier.

The theorem goes on to state that the rectangles do not need to be of the same width. Using the notation of Definition 6.8, let Δx_i denote the length of the i^{th} subinterval in a partition of $[a,b]$. Now let $||\Delta x||$ represent the length of the largest subinterval in the partition: that is, $||\Delta x||$ is the largest of all the Δx_i's. If $||\Delta x||$ is small, then $[a,b]$ must be partitioned into many subintervals, since all subintervals must have small lengths. "Taking the limit as $||\Delta x||$ goes to zero" implies that the number n of subintervals in the partition is growing to infinity, as the largest subinterval length is becoming arbitrarily small. We then interpret the expression

$$\lim_{||\Delta x||\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i$$

as "the limit of the sum of rectangles, where the width of each rectangle can be different but getting small, and the height of each rectangle is not necessarily determined by a particular rule." The following theorem states that, for a sufficiently nice function, we can use any of our three rules to find the area under $f(x)$ over $[a,b]$.

Theorem 6.12: Area and the Limit of Riemann Sums

Let $f(x)$ be a continuous function on the closed interval $[a,b]$ and let $S_L(n)$, $S_R(n)$ and $S_M(n)$ be the sums of equally spaced rectangles formed using the Left Hand Rule, Right Hand Rule, and Midpoint Rule, respectively. Then:

1. $\lim\limits_{n\to\infty} S_L(n) = \lim\limits_{n\to\infty} S_R(n) = \lim\limits_{n\to\infty} S_M(n) = \lim\limits_{n\to\infty} \sum_{i=1}^{n} f(c_i)\Delta x$

2. *The area under f on the interval $[a,b]$ is equal to* $\lim\limits_{n\to\infty} \sum_{i=1}^{n} f(c_i)\Delta x$.

3. *The area under f on the interval $[a,b]$ is equal to* $\lim\limits_{\|\Delta x\|\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i$.

We summarize what we have learned over the past few sections here.

- Knowing the "area under the curve" can be useful. One common example is: the area under a velocity curve is displacement.
- While we can approximate the area under a curve in many ways, we have focused on using rectangles whose heights can be determined using: the Left Hand Rule, the Right Hand Rule and the Midpoint Rule.
- Sums of rectangles of this type are called Riemann sums.
- The exact value of the area can be computed using the limit of a Riemann sum. We generally use one of the above methods as it makes the algebra simpler.

Exercises for Section 6.1

Exercise 6.1.1 *Suppose an object moves in a straight line so that its speed at time t is given by $v(t) = 2t+2$, and that at $t = 1$ the object is at position 5. Find the position of the object at $t = 2$.*

Exercise 6.1.2 *Suppose an object moves in a straight line so that its speed at time t is given by $v(t) = t^2+2$, and that at $t = 0$ the object is at position 5. Find the position of the object at $t = 2$.*

Exercise 6.1.3 *Find the area under $y = 2x$ between $x = 0$ and any positive value for x.*

Exercise 6.1.4 *Find the area under $y = 4x$ between $x = 0$ and any positive value for x.*

Exercise 6.1.5 *Find the area under $y = 4x$ between $x = 2$ and any positive value for x bigger than 2.*

Exercise 6.1.6 *Find the area under $y = 4x$ between any two positive values for x, say $a < b$.*

Exercise 6.1.7 *Let $f(x) = x^2 + 3x + 2$. Approximate the area under the curve between $x = 0$ and $x = 2$ using 4 rectangles and also using 8 rectangles.*

Exercise 6.1.8 *Let $f(x) = x^2 - 2x + 3$. Approximate the area under the curve between $x = 1$ and $x = 3$ using 4 rectangles.*

6.2 The Fundamental Theorem of Calculus

We begin by exploring an example. Suppose that an object moves in a straight line so that its speed is $3t$ at time t. How far does the object travel between time $t = a$ and time $t = b$? We don't assume that we know where the object is at time $t = 0$ or at any other time. It is certainly true that it is *somewhere,* so let's suppose that at $t = 0$ the position is k. Then we know that the position of the object at any time is $3t^2/2 + k$. This means that at time $t = a$ the position is $3a^2/2 + k$ and at time $t = b$ the position is $3b^2/2 + k$. Therefore the change in position is $3b^2/2 + k - (3a^2/2 + k) = 3b^2/2 - 3a^2/2$. Notice that the k drops out; this means that it doesn't matter that we don't know k, it doesn't even matter if we use the wrong k, we get the correct answer.

What about a second approach to this problem? We now want to approximate the change in position between time a and time b. We take the interval of time between a and b, divide it into n subintervals, and approximate the distance traveled during each. The starting time of subinterval number i is now $a + (i - 1)(b - a)/n$, which we abbreviate as t_{i-1}, so that $t_0 = a$, $t_1 = a + (b - a)/n$, and so on. The speed of the object is $f(t) = 3t$, and each subinterval is $(b - a)/n = \Delta t$ seconds long. The distance traveled during subinterval number i is approximately $f(t_{i-1})\Delta t$, and the total change in distance is approximately

$$f(t_0)\Delta t + f(t_1)\Delta t + \cdots + f(t_{n-1})\Delta t.$$

The exact change in position is the limit of this sum as n goes to infinity. We abbreviate this sum using **sigma notation**:

$$\sum_{i=0}^{n-1} f(t_i)\Delta t = f(t_0)\Delta t + f(t_1)\Delta t + \cdots + f(t_{n-1})\Delta t.$$

The notation on the left side of the equal sign uses a large capital sigma, a Greek letter, and the left side is an abbreviation for the right side. The answer we seek is

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t.$$

Since this must be the same as the answer we have already obtained, we know that

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t = \frac{3b^2}{2} - \frac{3a^2}{2}.$$

The significance of $3t^2/2$, into which we substitute $t = b$ and $t = a$, is of course that it is a function whose derivative is $f(t)$. As we have discussed, by the time we know that we want to compute

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t,$$

it no longer matters what $f(t)$ stands for—it could be a speed, or the height of a curve, or something else entirely. We know that the limit can be computed by finding any function with derivative $f(t)$, substituting a and b, and subtracting. We summarize this in a theorem. First, we introduce some new notation and terms.

We write

$$\int_a^b f(t)\,dt = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t$$

if the limit exists. That is, the left hand side means, or is an abbreviation for, the right hand side. The symbol $\int$ is called an **integral sign**, and the whole expression is read as "the integral of $f(t)$ from a to b." What we have learned is that this integral can be computed by finding a function, say $F(t)$, with the property that $F'(t) = f(t)$, and then computing $F(b) - F(a)$. The function $F(t)$ is called an **antiderivative** of $f(t)$. Now the theorem:

Theorem 6.13: Fundamental Theorem of Calculus

Suppose that $f(x)$ is continuous on the interval $[a,b]$. If $F(x)$ is any antiderivative of $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Let's rewrite this slightly:

$$\int_a^x f(t)\,dt = F(x) - F(a).$$

We've replaced the variable x by t and b by x. These are just different names for quantities, so the substitution doesn't change the meaning. It does make it easier to think of the two sides of the equation as functions. The expression

$$\int_a^x f(t)\,dt$$

is a function: plug in a value for x, get out some other value. The expression $F(x) - F(a)$ is of course also a function, and it has a nice property:

$$\frac{d}{dx}(F(x) - F(a)) = F'(x) = f(x),$$

since $F(a)$ is a constant and has derivative zero. In other words, by shifting our point of view slightly, we see that the odd looking function

$$G(x) = \int_a^x f(t)\,dt$$

has a derivative, and that in fact $G'(x) = f(x)$. This is really just a restatement of the Fundamental Theorem of Calculus, and indeed is often called the Fundamental Theorem of Calculus. To avoid confusion, some people call the two versions of the theorem "The Fundamental Theorem of Calculus, part I" and "The Fundamental Theorem of Calculus, part II", although unfortunately there is no universal agreement as to which is part I and which part II. Since it really is the same theorem, differently stated, some people simply call them both "The Fundamental Theorem of Calculus."

Theorem 6.14: Fundamental Theorem of Calculus

Suppose that $f(x)$ is continuous on the interval $[a,b]$ and let

$$G(x) = \int_a^x f(t)\,dt.$$

Then $G'(x) = f(x)$.

We have not really proved the Fundamental Theorem. In a nutshell, we gave the following argument to justify it: Suppose we want to know the value of

$$\int_a^b f(t)\,dt = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\Delta t.$$

We can interpret the right hand side as the distance traveled by an object whose speed is given by $f(t)$. We know another way to compute the answer to such a problem: find the position of the object by finding an antiderivative of $f(t)$, then substitute $t = a$ and $t = b$ and subtract to find the distance traveled. This must be the answer to the original problem as well, even if $f(t)$ does not represent a speed.

What's wrong with this? In some sense, nothing. As a practical matter it is a very convincing argument, because our understanding of the relationship between speed and distance seems to be quite solid. From the point of view of mathematics, however, it is unsatisfactory to justify a purely mathematical relationship by appealing to our understanding of the physical universe, which could, however unlikely it is in this case, be wrong.

A complete proof is a bit too involved to include here, but we will indicate how it goes. First, if we can prove the second version of the Fundamental Theorem, Theorem 6.14, then we can prove the first version from that:

Proof. Proof of Theorem 6.13.

We know from Theorem 6.14 that

$$G(x) = \int_a^x f(t)\,dt$$

is an antiderivative of $f(x)$, and therefore any antiderivative $F(x)$ of $f(x)$ is of the form $F(x) = G(x) + k$. Then

$$\begin{aligned} F(b) - F(a) = G(b) + k - (G(a) + k) &= G(b) - G(a) \\ &= \int_a^b f(t)\,dt - \int_a^a f(t)\,dt. \end{aligned}$$

It is not hard to see that $\int_a^a f(t)\,dt = 0$, so this means that

$$F(b) - F(a) = \int_a^b f(t)\,dt,$$

which is exactly what Theorem 6.13 says. ♣

So the real job is to prove Theorem 6.14. We will sketch the proof, using some facts that we do not prove. First, the following identity is true of integrals:

$$\int_a^b f(t)\,dt = \int_a^c f(t)\,dt + \int_c^b f(t)\,dt.$$

This can be proved directly from the definition of the integral, that is, using the limits of sums. It is quite easy to see that it must be true by thinking of either of the two applications of integrals that we have seen. It turns out that the identity is true no matter what c is, but it is easiest to think about the meaning when $a \le c \le b$.

First, if $f(t)$ represents a speed, then we know that the three integrals represent the distance traveled between time a and time b; the distance traveled between time a and time c; and the distance traveled between time c and time b. Clearly the sum of the latter two is equal to the first of these.

Second, if $f(t)$ represents the height of a curve, the three integrals represent the area under the curve between a and b; the area under the curve between a and c; and the area under the curve between c and b. Again it is clear from the geometry that the first is equal to the sum of the second and third.

Proof. Proof of Theorem 6.14.

We want to compute $G'(x)$, so we start with the definition of the derivative in terms of a limit:

$$\begin{aligned} G'(x) &= \lim_{\Delta x \to 0} \frac{G(x+\Delta x)-G(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left(\int_a^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt\right) \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left(\int_a^x f(t)\,dt + \int_x^{x+\Delta x} f(t)\,dt - \int_a^x f(t)\,dt\right) \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\int_x^{x+\Delta x} f(t)\,dt. \end{aligned}$$

Now we need to know something about

$$\int_x^{x+\Delta x} f(t)\,dt$$

when Δx is small; in fact, it is very close to $\Delta x f(x)$, but we will not prove this. Once again, it is easy to believe this is true by thinking of our two applications: The integral

$$\int_x^{x+\Delta x} f(t)\,dt$$

can be interpreted as the distance traveled by an object over a very short interval of time. Over a sufficiently short period of time, the speed of the object will not change very much, so the distance traveled will be approximately the length of time multiplied by the speed at the beginning of the interval, namely, $\Delta x f(x)$. Alternately, the integral may be interpreted as the area under the curve between x and $x+\Delta x$. When Δx is very small, this will be very close to the area of the rectangle with base Δx and height $f(x)$; again this is $\Delta x f(x)$. If we accept this, we may proceed:

$$\lim_{\Delta x \to 0} \frac{1}{\Delta x}\int_x^{x+\Delta x} f(t)\,dt = \lim_{\Delta x \to 0} \frac{\Delta x f(x)}{\Delta x} = f(x),$$

which is what we wanted to show. ♣

It is still true that we are depending on an interpretation of the integral to justify the argument, but we have isolated this part of the argument into two facts that are not too hard to prove. Once the last reference to interpretation has been removed from the proofs of these facts, we will have a real proof of the Fundamental Theorem.

Now we know that to solve certain kinds of problems, those that lead to a sum of a certain form, we "merely" find an antiderivative and substitute two values and subtract. Unfortunately, finding antiderivatives can be quite difficult. While there are a small number of rules that allow us to compute the derivative of any common function, there are no such rules for antiderivatives. There are some techniques that frequently prove useful, but we will never be able to reduce the problem to a completely mechanical process.

Due to the close relationship between an integral and an antiderivative, the integral sign is also used to mean "antiderivative". You can tell which is intended by whether the limits of integration are included:

$$\int_1^2 x^2\,dx$$

is an ordinary integral, also called a **definite integral**, because it has a definite value, namely

$$\int_1^2 x^2\,dx = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

We use

$$\int x^2\,dx$$

to denote the antiderivative of x^2, also called an **indefinite integral**. So this is evaluated as

$$\int x^2\,dx = \frac{x^3}{3} + C.$$

It is customary to include the constant C to indicate that there are really an infinite number of antiderivatives. We do not need this C to compute definite integrals, but in other circumstances we will need to remember that the C is there, so it is best to get into the habit of writing the C. When we compute a definite integral, we first find an antiderivative and then substitute. It is convenient to first display the antiderivative and then do the substitution; we need a notation indicating that the substitution is yet to be done. A typical solution would look like this:

$$\int_1^2 x^2\,dx = \left.\frac{x^3}{3}\right|_1^2 = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

The vertical line with subscript and superscript is used to indicate the operation "substitute and subtract" that is needed to finish the evaluation.

We seem to have found a pattern. When attempting to solve a previous question, we found the antiderivative of x^2 to be $x^3/3+c$ (as it was when solving the indefinite integral). Likewise, when we first began, we were trying to determine a position based on velocity, and $3t$ gave rise to $3t^2/2+k$.

As will be formalized later, we see that in these cases, the power is increased to $n+1$, but we also divide through by this factor, $n+1$. So x becomes $x^2/2$, x^2 becomes $x^3/3$, and x^3 will become $x^4/4$.

Now we will also try with negative and fraction values in the following example.

Example 6.15: Fundamental Theorem of Calculus

Evaluate $\displaystyle\int_1^4 x^3 + \sqrt{x} + \frac{1}{x^2}\,dx.$

Solution.

$$\begin{aligned}
\int_1^4 x^3+\sqrt{x}+\frac{1}{x^2}\,dx &= \left.\frac{x^4}{4}+\frac{2x^{3/2}}{3}-x^{-1}\right|_1^4 \\
&= \left(\frac{(4)^4}{4}+\frac{2(4)^{3/2}}{3}-4^{-1}\right) \\
&\quad -\left(\frac{(1)^4}{4}+\frac{2(1)^{3/2}}{3}-1^{-1}\right) \\
&= \frac{415}{6}
\end{aligned}$$

♣

Properties of Definite Integrals

Some properties are as follows:

Order of limits matters: $\displaystyle\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$

If interval is empty, integral is zero: $\displaystyle\int_a^a f(x)\,dx = 0$

Constant Multiple Rule: $\displaystyle\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$

Sum/Difference Rule: $\displaystyle\int_a^b f(x)\pm g(x)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$

Can split up interval $[a,b]=[a,c]\cup[c,b]$: $\displaystyle\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$

The variable does not matter!: $\displaystyle\int_a^b f(x)\,dx = \int_a^b f(t)\,dt$

The reason for the last property is that a definite integral is a *number*, not a function, so the variable is just a placeholder that won't appear in the final answer.

Some additional properties are *comparison* types of properties.

Comparison Properties of Definite Integrals

$$\text{If } f(x) \geq 0 \text{ for } x \in [a,b], \text{ then:} \qquad \int_a^b f(x)\,dx \geq 0.$$

$$\text{If } f(x) \geq g(x) \text{ for } x \in [a,b], \text{ then:} \qquad \int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx.$$

$$\text{If } m \leq f(x) \leq M \text{ for } x \in [a,b], \text{ then:} \qquad m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

Example 6.16: Properties of Definite Integrals

Suppose $\int_a^b f(x)\,dx = 7$ *and* $\int_a^b g(x)\,dx = 3$. *Find:*

1. $\int_a^b 2f(x) - 3g(x)\,dx.$
2. $\int_b^a 2g(x)\,dx.$
3. $\int_a^a f(x) \cdot g(x)\,dx.$
4. $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$

Solution.

1. $\int_a^b 2f(x) - 3g(x)\,dx = 2\int_a^b f(x)\,dx - 3\int_a^b g(x)\,dx = 2(7) - 3(3) = 5.$
2. $\int_b^a 2g(x)\,dx = -2\int_a^b g(x)\,dx = -2(3) = -6.$
3. $\int_a^a f(x) \cdot g(x)\,dx = 0.$
4. $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = \int_a^b f(x)\,dx = 7.$

♣

We next evaluate a definite integral using three different techniques.

Example 6.17: Three Different Techniques

Evaluate $\int_0^2 x+1\,dx$ *by*

1. *Using FTC II (the shortcut)*
2. *Using the definition of a definite integral (the limit sum definition)*
3. *Interpreting the problem in terms of areas (graphically)*

Solution. 1. The shortcut (FTC II) is the method of choice as it is the fastest. Integrating and using the '*top minus bottom*' rule we have:

$$\begin{aligned}\int_0^2 x+1\,dx &= \left.\frac{x^2}{2}+x\right|_0^2 \\ &= \left[\frac{2^2}{2}+2\right]-\left[\frac{0^2}{2}+0\right]=4.\end{aligned}$$

2. We now use the definition of a definite integral. We divide the interval $[0,2]$ into n subintervals of equal width Δx, and from each interval choose a point x_i^*. Using the formulas

$$\Delta x=\frac{b-a}{n} \qquad \text{and} \qquad x_i=a+i\Delta x,$$

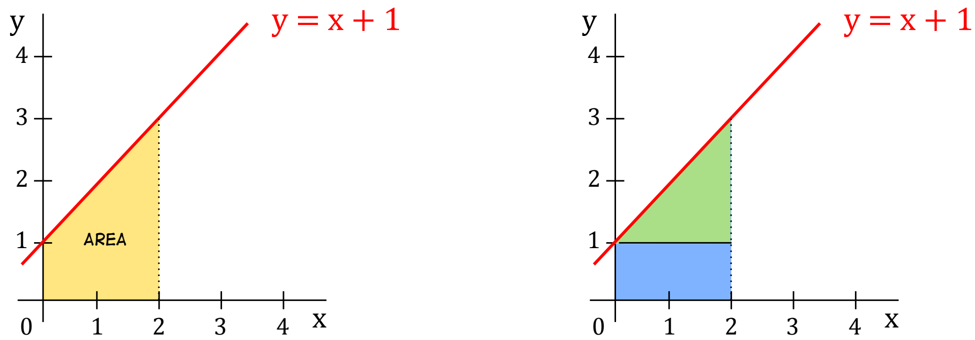
we have

$$\Delta x=\frac{2}{n} \qquad \text{and} \qquad x_i=0+i\Delta x=\frac{2i}{n}.$$

Then taking x_i^*'s as right endpoints for convenience (so that $x_i^*=x_i$), we have:

$$\begin{aligned}\int_0^2 x+1\,dx &= \lim_{n\to\infty}\sum_{i=1}^n f(x_i^*)\Delta x \\ &= \lim_{n\to\infty}\sum_{i=1}^n f\left(\frac{2i}{n}\right)\frac{2}{n} \\ &= \lim_{n\to\infty}\sum_{i=1}^n \left(\frac{2i}{n}+1\right)\frac{2}{n} \\ &= \lim_{n\to\infty}\sum_{i=1}^n \left(\frac{4i}{n^2}+\frac{2}{n}\right) \\ &= \lim_{n\to\infty}\left(\sum_{i=1}^n \frac{4i}{n^2}+\sum_{i=1}^n \frac{2}{n}\right) \\ &= \lim_{n\to\infty}\left(\frac{4}{n^2}\sum_{i=1}^n i+\frac{2}{n}\sum_{i=1}^n 1\right) \\ &= \lim_{n\to\infty}\left(\frac{4}{n^2}\frac{n(n+1)}{2}+\frac{2}{n}n\right) \\ &= \lim_{n\to\infty}\left(2+\frac{2}{n}+2\right) \\ &= 4.\end{aligned}$$

3. Finally, let's evaluate the net area under $x+1$ from 0 to 2.



Thus, the area is the sum of the areas of a rectangle and a triangle. Hence,

$$\begin{aligned}\int_0^2 x+1\,dx &= \text{Net Area} \\ &= \text{Area of rectangle} + \text{Area of triangle} \\ &= (2)(1)+\frac{1}{2}(2)(2) \\ &= 4.\end{aligned}$$

♣

We next apply FTC to differentiate a function.

Example 6.18: Using FTC

Differentiate the following function:

$$g(x)=\int_{-2}^{x}\cos(1+5t)\sin t\,dt.$$

Solution. We simply apply the Fundamental Theorem of Calculus directly to get:

$$g'(x)=\cos(1+5x)\sin x.$$

♣

Using the Chain Rule we can derive a formula for some more complicated problems. We have:

$$\frac{d}{dx}\int_a^{v(x)} f(t)\,dt = f(v(x))\cdot v'(x).$$

Now what if the upper limit is constant and the lower limit is a function of x? Then we interchange the limits and add a minus sign to get:

$$\frac{d}{dx}\int_{u(x)}^{a} f(t)\,dt = -\frac{d}{dx}\int_a^{u(x)} f(t)\,dt = -f(u(x))\cdot u'(x).$$

Combining these two we can get a formula where both limits are a function of x. We break up the integral as follows:

$$\int_{u(x)}^{v(x)} f(t)\,dt = \int_{u(x)}^{a} f(t)\,dt + \int_{a}^{v(x)} f(t)\,dt.$$

We just need to make sure $f(a)$ exists after we break up the integral. Then differentiating and using the above two formulas gives:

FTC I + Chain Rule:

$$\frac{d}{dx}\int_{u(x)}^{v(x)} f(t)\,dt = f(v(x))v'(x) - f(u(x))u'(x)$$

Many textbooks do not show this formula and instead to solve these types of problems will use FTC I along with the tricks we used to derive the formula above. Either method is perfectly fine.

Example 6.19: FTC I + Chain Rule

Differentiate the following integral:

$$\int_{10x}^{x^2} t^3\sin(1+t)\,dt.$$

Solution. We will use the formula above. We have $f(t) = t^3\sin(1+t)$, $u(x) = 10x$ and $v(x) = x^2$. Then $u'(x) = 10$ and $v'(x) = 2x$. Thus,

$$\begin{aligned} \frac{d}{dx}\int_{10x}^{x^2} t^3\sin(1+t)\,dt &= (x^2)^3\sin(1+(x^2))(2x) - (10x)^3\sin(1+(10x))(10) \\ &= 2x^7\sin(1+x^2) - 10000x^3\sin(1+10x) \end{aligned}$$

♣

Example 6.20: FTC I + Chain Rule

Differentiate the following integral with respect to x:

$$\int_{x^3}^{2x} 1+\cos t\,dt$$

Solution. Using the formula we have:

$$\frac{d}{dx}\int_{x^3}^{2x} 1+\cos t\,dt = (1+\cos(2x))(2) - (1+\cos(x^3))(3x^2).$$

♣

Exercises for Section 6.2

Exercise 6.2.1 *Evaluate* $\displaystyle\int_1^4 t^2 + 3t\,dt$

Exercise 6.2.2 *Evaluate* $\displaystyle\int_0^\pi \sin t\,dt$

Exercise 6.2.3 *Evaluate* $\displaystyle\int_1^{10} \frac{1}{x}\,dx$

Exercise 6.2.4 *Evaluate* $\displaystyle\int_0^5 e^x\,dx$

Exercise 6.2.5 *Evaluate* $\displaystyle\int_0^3 x^3\,dx$

Exercise 6.2.6 *Evaluate* $\displaystyle\int_1^2 x^5\,dx$

Exercise 6.2.7 *Find the derivative of* $\displaystyle G(x) = \int_1^x t^2 - 3t\,dt$

Exercise 6.2.8 *Find the derivative of* $\displaystyle G(x) = \int_1^{x^2} t^2 - 3t\,dt$

Exercise 6.2.9 *Find the derivative of* $\displaystyle G(x) = \int_1^x e^{t^2}\,dt$

Exercise 6.2.10 *Find the derivative of* $\displaystyle G(x) = \int_1^{x^2} e^{t^2}\,dt$

Exercise 6.2.11 *Find the derivative of* $\displaystyle G(x) = \int_1^x \tan(t^2)\,dt$

Exercise 6.2.12 *Find the derivative of* $\displaystyle G(x) = \int_{10x}^{x^2} \tan(t^2)\,dt$

Exercise 6.2.13 *Suppose* $\int_1^4 f(x)\,dx = 2$ *and* $\int_1^4 g(x)\,dx = 7$. *Find* $\int_1^4 (5f(x) + 3g(x))\,dx$ *and* $\int_1^4 (6 - 2f(x))\,dx$.

Exercise 6.2.14 *Suppose* $\int_{-2}^5 f(x)\,dx = 3$ *and* $\int_1^5 f(x)\,dx = -2$. *Find* $\int_{-2}^1 f(x)\,dx$.

Exercise 6.2.15 *If* f *is continuous on* $[a,b]$, *we define the average of* $f(x)$ *on* $[a,b]$ *to be*

$$avg_{[a,b]}(f) = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

(a) What is the average of $\sqrt{x}$ *on the interval* $[0,1]$?

(b) If the average of $f(x)$ *on* $[0,2]$ *and on* $[2,5]$ *are 6 and 4 respectively, then what is its average on* $[0,5]$*?*

6.3 Indefinite Integrals

In this section we focus on computing indefinite integrals. The process of finding the indefinite integral is called **integration** (or **integrating** $f(x)$).

Example 6.21: Indefinite Integral

Evaluate the following indefinite integral:

$$\int x^5 + 3x - 2\,dx.$$

Solution. Since this is asking for the most general anti-derivative we have:

$$\int x^5 + 3x - 2\,dx = \frac{x^6}{6} + \frac{3x^2}{2} - 2x + C$$

where C is a constant. ♣

Common mistakes: One habit students make with integrals is to *drop the dx* at the end of the integral. This is required! Think of the integral as a set of parenthesis. Both are required so it is clear where the integrand ends and what variable you are integrating with respect to.

Another common mistake is to *forget the* $+C$ for indefinite integrals.

Note that we don't have properties to deal with products or quotients of functions, that is,

$$\int f(x) \cdot g(x)\,dx \neq \int f(x)\,dx \int g(x)\,dx.$$

$$\int \frac{f(x)}{g(x)}\,dx \neq \frac{\int f(x)\,dx}{\int g(x)\,dx}.$$

With derivatives, we had the product and quotient rules to deal with these cases. For integrals, we have no such rules, but we will learn a variety of different techniques to deal with these cases.

The following integral rules can be proved by taking the derivative of the functions on the right side.

Integral Rules

Some properties and rules to know:

$$\text{Constant Rule:} \quad \int k\,dx = kx + C.$$

$$\text{Constant Multiple Rule:} \quad \int kf(x)\,dx = k\int f(x)\,dx, \quad k \text{ is constant.}$$

$$\text{Sum/Difference Rule:} \quad \int f(x) \pm g(x)\,dx = \int f(x)\,dx \pm \int g(x)\,dx.$$

$$\text{Power Rule:} \quad \int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq -1.$$

$$\text{Log Rule:} \quad \int \frac{1}{x}\,dx = \ln|x| + C, \quad x \neq 0.$$

$$\text{Exponent Rule:} \quad \int a^{kx} = \frac{a^{kx}}{k\ln a} + C, \quad x \neq 0.$$

$$\text{Sine Rule:} \quad \int \sin x\,dx = -\cos x + C.$$

$$\text{Cosine Rule:} \quad \int \cos x\,dx = \sin x + C.$$

Example 6.22: Indefinite Integral

If $f'(x) = x^4 + 2x - 8\sin x$ then what is $f(x)$?

Solution. The answer is:

$$\begin{aligned} f(x) = \int f'(x)\,dx &= \int \left(x^4 + 2x - 8\sin x\right)\,dx \\ &= \int x^4\,dx + 2\int x\,dx - 8\int \sin x\,dx \\ &= \frac{x^5}{5} + x^2 + 8\cos x + C, \end{aligned}$$

where C is a constant. ♣

Example 6.23: Indefinite Integral

Find the general indefinite integral of $\int 3x^2\,dx$.

Solution.

$$\begin{aligned}\int 3x^2\,dx &= 3\int x^2\,dx \\ &= 3\frac{x^3}{3}+C \\ &= x^3+C\end{aligned}$$

♣

Example 6.24: Indefinite Integral

Find the general indefinite integral of $\int \frac{2}{\sqrt{x}}\,dx$.

Solution.

$$\begin{aligned}\int \frac{2}{\sqrt{x}}\,dx &= 2\int x^{-\frac{1}{2}}\,dx \\ &= 2\frac{x^{-\frac{1}{2}+1}}{-\frac{1}{2}+1}+C \\ &= 4\sqrt{x}+C\end{aligned}$$

♣

Example 6.25: Indefinite Integral

Find the general indefinite integral of $\int \left(\frac{1}{x}+e^{7x}+x^{\pi}+7\right)\,dx$.

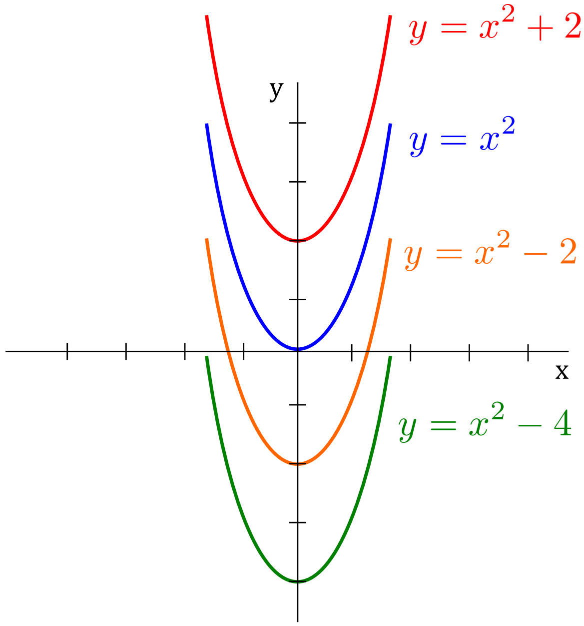
Solution.

$$\begin{aligned}\int \left(\frac{1}{x}+e^{7x}+x^{\pi}+7\right)\,dx &= \int \frac{1}{x}\,dx+\int e^{7x}\,dx+\int x^{\pi}\,dx+\int 7\,dx \\ &= \ln|x|+\frac{1}{7}e^{7x}+\frac{x^{\pi+1}}{\pi+1}+7x+C\end{aligned}$$

♣

Differential Equations

An equation involving derivatives where we want to solve for the original function is called a **differential equation**. For example, $f'(x)=2x$ is a differential equation with general solution $f(x)=x^2+C$. Some solutions (i.e., particular values of C) are shown below.



As seen with integral curves, we may have an infinite family of solutions satisfying the differential equation. However, if we were given a point (called an *initial value*) on the curve then we could determine $f(x)$ completely. Such a problem is known as an *initial value problem*.

Example 6.26: Initial Value Problem

If $f'(x) = 2x$ and $f(0) = 2$ then determine $f(x)$.

Solution. As previously stated, we have a solution of:

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

But $f(0) = 2$ implies:

$$2 = 0^2 + C \quad \rightarrow \quad C = 2.$$

Therefore, $f(x) = x^2 + 2$ is the solution to the initial value problem. ♣

Exercises for Section 6.3

Find the antiderivatives of the functions:

Exercise 6.3.1 $8\sqrt{x}$

Exercise 6.3.2 $3t^2 + 1$

Exercise 6.3.3 $4/\sqrt{x}$

Exercise 6.3.4 $2/z^2$

Exercise 6.3.5 $7s^{-1}$

Exercise 6.3.6 $(5x+1)^2$

Exercise 6.3.7 $(x-6)^2$

Exercise 6.3.8 $x^{3/2}$

Exercise 6.3.9 $\dfrac{2}{x\sqrt{x}}$

Exercise 6.3.10 $|2t-4|$

7. Techniques of Integration

Over the next few sections we examine some techniques that are frequently successful when seeking antiderivatives of functions.

7.1 Substitution Rule

Needless to say, most integration problems we will encounter will not be so simple. That is to say we will require more than the basic integration rules we have seen. Here's a slightly more complicated example: Find

$$\int 2x\cos(x^2)\,dx.$$

This is not a "simple" derivative, but a little thought reveals that it must have come from an application of the chain rule. Multiplied on the "outside" is $2x$, which is the derivative of the "inside" function x^2. Checking:

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

so

$$\int 2x\cos(x^2)\,dx = \sin(x^2) + C.$$

To summarize: If we suspect that a given function is the derivative of another via the chain rule, we let u denote a likely candidate for the inner function, then translate the given function so that it is written entirely in terms of u, with no x remaining in the expression. If we can integrate this new function of u, then the antiderivative of the original function is obtained by replacing u by the equivalent expression in x.

Theorem 7.1: Substitution Rule for Indefinite Integrals

If $u = g(x)$ is a differentiable function whose range is an interval I and f is continuous on I, then

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du.$$

Even in simple cases you may prefer to use this mechanical procedure, since it often helps to avoid silly mistakes. For example, consider again this simple problem:

$$\int 2x\cos(x^2)\,dx.$$

Let $u = x^2$, then $du/dx = 2x$ or $du = 2x\,dx$. Since we have exactly $2x\,dx$ in the original integral, we can replace it by du:

$$\int 2x\cos(x^2)\,dx = \int \cos u\,du = \sin u + C = \sin(x^2) + C.$$

This is not the only way to do the algebra, and typically there are many paths to the correct answer. Another possibility, for example, is: Since $du/dx = 2x$, $dx = du/2x$, and then the integral becomes

$$\int 2x\cos(x^2)\,dx = \int 2x\cos u\,\frac{du}{2x} = \int \cos u\,du.$$

The important thing to remember is that you must eliminate all instances of the original variable x.

Example 7.2: Substitution Rule

Evaluate $\int (ax+b)^n\,dx$, assuming a, b are constants, $a \neq 0$, and n is a positive integer.

Solution. We let $u = ax+b$ so $du = a\,dx$ or $dx = du/a$. Then

$$\int (ax+b)^n\,dx = \int \frac{1}{a}u^n\,du = \frac{1}{a(n+1)}u^{n+1} + C = \frac{1}{a(n+1)}(ax+b)^{n+1} + C.$$

♣

Example 7.3: Substitution Rule

Evaluate $\int \sin(ax+b)\,dx$, assuming that a and b are constants and $a \neq 0$.

Solution. Again we let $u = ax+b$ so $du = a\,dx$ or $dx = du/a$. Then

$$\int \sin(ax+b)\,dx = \int \frac{1}{a}\sin u\,du = \frac{1}{a}(-\cos u) + C = -\frac{1}{a}\cos(ax+b) + C.$$

♣

Strategy for Substitution Rule

A general strategy to follow is:

1. Choose a possible $u = u(x)$. **Tip:** Choose a substitution u so that its derivate also appears in the integral (up to a constant).
2. Calculate $du = u'(x)\,dx$.
3. Either replace $u'(x)\,dx$ by du, or replace dx by $\dfrac{du}{u'(x)}$, and cancel.
4. Write the rest of the integrand in terms of u. If this is not possible, the substitution will not work: You must go back to step 1.
5. Find the indefinite integral. (Again, if this is not possible, try a different substitution, or a different method).
6. Rewrite the result in terms of x.

Example 7.4: Substitution

Evaluate the following integral: $\int \frac{2x}{\sqrt{1-4x^2}}\,dx.$

Solution. We try the substitution:

$$u = 1-4x^2.$$

Then,

$$du = -8x\,dx$$

In the numerator we have $2x\,dx$, so rewriting the differential gives:

$$-\frac{1}{4}du = 2x\,dx.$$

Then the integral is:

$$\begin{aligned}
\int \frac{2x}{\sqrt{1-4x^2}}\,dx &= \int \left(1-4x^2\right)^{-1/2}(2x\,dx) \\
&= \int u^{-1/2}\left(-\frac{1}{4}du\right) \\
&= \left(\frac{-1}{4}\right)\frac{u^{1/2}}{1/2}+C \\
&= -\frac{\sqrt{1-4x^2}}{2}+C
\end{aligned}$$

♣

Example 7.5: Substitution

Evaluate the following integral: $\int \cos x(\sin x)^5\,dx.$

Solution. In this question we will let $u = \sin x$. Then,

$$du = \cos x\,dx.$$

Thus, the integral becomes:

$$\begin{aligned}
\int \cos x(\sin x)^5\,dx &= \int u^5\,du \\
&= \frac{u^6}{6}+C \\
&= \frac{(\sin x)^6}{6}+C
\end{aligned}$$

♣

Example 7.6: Substitution

Evaluate the following integral: $\displaystyle\int \frac{\cos(\sqrt{x})}{\sqrt{x}}\,dx.$

Solution. We use the substitution:

$$u = x^{1/2}.$$

Then,

$$du = \frac{1}{2}x^{-1/2}dx.$$

Rewriting the differential we get:

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

The integral becomes:

$$\begin{aligned}
\int \frac{\cos(\sqrt{x})}{\sqrt{x}}\,dx &= 2\int \cos u\,du \\
&= 2\sin u + C \\
&= 2\sin(\sqrt{x}) + C
\end{aligned}$$

♣

Example 7.7: Substitution

Evaluate the following integral: $\displaystyle\int 2x^3\sqrt{x^2+1}\,dx.$

Solution. This problem is a little bit different than the previous ones. It makes sense to let:

$$u = x^2 + 1,$$

then

$$du = 2x\,dx.$$

Making this substitution gives:

$$\begin{aligned}
\int 2x^3\sqrt{x^2+1}\,dx &= \int x^2\sqrt{x^2+1}(2x)\,dx \\
&= \int x^2u^{1/2}\,du
\end{aligned}$$

This is a problem because our integrals can't have a mixture of two variables in them. Usually this means we chose our u incorrectly. However, in this case we can eliminate the remaining x's from our integral by using:

$$u = x^2 + 1 \quad \to \quad x^2 = u - 1.$$

We get:

$$\begin{aligned}\int x^2 u^{1/2}\, du &= \int (u-1)u^{1/2}\, du \\ &= \int u^{3/2} - u^{1/2}\, du \\ &= \frac{2}{5}u^{5/2} - \frac{2}{3}u^{3/2} + C \\ &= \frac{2}{5}(x^2+1)^{5/2} - \frac{2}{3}(x^2+1)^{3/2} + C\end{aligned}$$

♣

The next example shows how to use the Substitution Rule when dealing with definite integrals.

Example 7.8: Substitution Rule

Evaluate $\int_2^4 x\sin(x^2)\, dx$.

Solution. First we compute the antiderivative, then evaluate the integral. Let $u = x^2$, so $du = 2x\,dx$ or $x\,dx = du/2$. Then

$$\int x\sin(x^2)\, dx = \int \frac{1}{2}\sin u\, du = \frac{1}{2}(-\cos u) + C = -\frac{1}{2}\cos(x^2) + C.$$

Now

$$\int_2^4 x\sin(x^2)\, dx = -\frac{1}{2}\cos(x^2)\Big|_2^4 = -\frac{1}{2}\cos(16) + \frac{1}{2}\cos(4).$$

A somewhat neater alternative to this method is to change the original limits to match the variable u. Since $u = x^2$, when $x = 2$, $u = 4$, and when $x = 4$, $u = 16$. So we can do this:

$$\int_2^4 x\sin(x^2)\, dx = \int_4^{16} \frac{1}{2}\sin u\, du = -\frac{1}{2}(\cos u)\Big|_4^{16} = -\frac{1}{2}\cos(16) + \frac{1}{2}\cos(4).$$

An incorrect, and dangerous, alternative is something like this:

$$\int_2^4 x\sin(x^2)\, dx = \int_2^4 \frac{1}{2}\sin u\, du = -\frac{1}{2}\cos(u)\Big|_2^4 = -\frac{1}{2}\cos(x^2)\Big|_2^4 = -\frac{1}{2}\cos(16) + \frac{1}{2}\cos(4).$$

This is incorrect because $\int_2^4 \frac{1}{2}\sin u\, du$ means that u takes on values between 2 and 4, which is wrong. It is dangerous, because it is very easy to get to the point $-\frac{1}{2}\cos(u)\Big|_2^4$ and forget to substitute x^2 back in for

u, thus getting the incorrect answer $-\frac{1}{2}\cos(4)+\frac{1}{2}\cos(2)$. An acceptable alternative is something like:

$$\int_2^4 x\sin(x^2)\,dx = \int_{x=2}^{x=4} \frac{1}{2}\sin u\,du = -\frac{1}{2}\cos(u)\Big|_{x=2}^{x=4} = -\frac{1}{2}\cos(x^2)\Big|_2^4 = -\frac{\cos(16)}{2}+\frac{\cos(4)}{2}.$$

♣

To summarize, we have the following.

Theorem 7.9: Substitution Rule for Definite Integrals

If g' is continuous on $[a,b]$ and f is continuous on the range of $u=g(x)$, then

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du.$$

Example 7.10: Substitution Rule

Evaluate $\displaystyle\int_{1/4}^{1/2} \frac{\cos(\pi t)}{\sin^2(\pi t)}\,dt$.

Solution. Let $u=\sin(\pi t)$ so $du=\pi\cos(\pi t)\,dt$ or $du/\pi=\cos(\pi t)\,dt$. We change the limits to $\sin(\pi/4)=\sqrt{2}/2$ and $\sin(\pi/2)=1$. Then

$$\int_{1/4}^{1/2} \frac{\cos(\pi t)}{\sin^2(\pi t)}\,dt = \int_{\sqrt{2}/2}^{1} \frac{1}{\pi}\frac{1}{u^2}\,du = \int_{\sqrt{2}/2}^{1} \frac{1}{\pi}u^{-2}\,du = \frac{1}{\pi}\frac{u^{-1}}{-1}\Big|_{\sqrt{2}/2}^{1} = -\frac{1}{\pi}+\frac{\sqrt{2}}{\pi}.$$

♣

Exercises for Section 7.1

Find the following indefinite and definite integrals.

Exercise 7.1.1 $\displaystyle\int (1-t)^9\,dt$

Exercise 7.1.2 $\displaystyle\int (x^2+1)^2\,dx$

Exercise 7.1.3 $\displaystyle\int x(x^2+1)^{100}\,dx$

Exercise 7.1.4 $\displaystyle\int \frac{1}{\sqrt[3]{1-5t}}\,dt$

Exercise 7.1.5 $\displaystyle\int \sin^3 x \cos x\,dx$

Exercise 7.1.6 $\displaystyle\int x\sqrt{100-x^2}\,dx$

Exercise 7.1.7 $\displaystyle\int \frac{x^2}{\sqrt{1-x^3}}\,dx$

Exercise 7.1.8 $\displaystyle\int \cos(\pi t)\cos\big(\sin(\pi t)\big)\,dt$

Exercise 7.1.9 $\displaystyle\int \frac{\sin x}{\cos^3 x}\,dx$

Exercise 7.1.10 $\displaystyle\int \tan x\,dx$

Exercise 7.1.11 $\displaystyle\int_0^{\pi} \sin^5(3x)\cos(3x)\,dx$

Exercise 7.1.12 $\displaystyle\int \sec^2 x \tan x\,dx$

Exercise 7.1.13 $\displaystyle\int_0^{\sqrt{\pi}/2} x\sec^2(x^2)\tan(x^2)\,dx$

Exercise 7.1.14 $\displaystyle\int \frac{\sin(\tan x)}{\cos^2 x}\,dx$

Exercise 7.1.15 $\displaystyle\int_3^4 \frac{1}{(3x-7)^2}\,dx$

Exercise 7.1.16 $\displaystyle\int_0^{\pi/6} (\cos^2 x - \sin^2 x)\,dx$

Exercise 7.1.17 $\displaystyle\int \frac{6x}{(x^2-7)^{1/9}}\,dx$

Exercise 7.1.18 $\displaystyle\int_{-1}^{1} (2x^3-1)(x^4-2x)^6\,dx$

Exercise 7.1.19 $\displaystyle\int_{-1}^{1} \sin^7 x\,dx$

Exercise 7.1.20 $\displaystyle\int f(x)f'(x)\,dx$

7.2 Powers of Trigonometric Functions

Functions consisting of powers of the sine and cosine can be integrated by using substitution and trigonometric identities. These can sometimes be tedious, but the technique is straightforward. A similar technique is applicable to powers of secant and tangent (and also cosecant and cotangent, not discussed here).

The trigonometric substitutions we will focus on in this section are summarized in the table below:

Substitution	$u=\sin x$	$u=\cos x$	$u=\tan x$	$u=\sec x$
Derivative	$du=\cos x\,dx$	$du=-\sin x\,dx$	$du=\sec^2 x\,dx$	$du=\sec x\tan x\,dx$

An example will suffice to explain the approach.

Example 7.11: Odd Power of Sine

Evaluate $\int \sin^5 x\,dx$.

Solution. Rewrite the function:

$$\begin{aligned}\int \sin^5 x\,dx &= \int \sin x \sin^4 x\,dx\\ &= \int \sin x(\sin^2 x)^2\,dx\\ &= \int \sin x(1-\cos^2 x)^2\,dx.\end{aligned}$$

Now use $u=\cos x$, $du=-\sin x\,dx$:

$$\begin{aligned}\int \sin x(1-\cos^2 x)^2\,dx &= \int -(1-u^2)^2\,du\\ &= \int -(1-2u^2+u^4)\,du\\ &= \int 1+2u^2-u^4\,du\\ &= -u+\frac{2}{3}u^3-\frac{1}{5}u^5+C\\ &= -\cos x+\frac{2}{3}\cos^3 x-\frac{1}{5}\cos^5 x+C.\end{aligned}$$

♣

Observe that by taking the substitution $u=\cos x$ in the last example, we ended up with an even power of sine from which we can use the formula $\sin^2 x+\cos^2 x=1$ to replace any remaining sines. We then ended up with a polynomial in u in which we could expand and integrate quite easily.

This technique works for products of powers of sine and cosine. We summarize it below.

Products of Sine and Cosine

When evaluating $\int \sin^m x \cos^n x\, dx$:

1. **The power of sine is odd (m odd):**
(a) Use $u = \cos x$ and $du = -\sin x\, dx$.
(b) Replace dx using (a), thus cancelling one power of $\sin x$ by the substitution of du, and be left with an even number of sine powers.
(c) Use $\sin^2 x = 1 - \cos^2 x$ $(= 1 - u^2)$ to replace the leftover sines.

2. **The power of cosine is odd (n odd):**
(a) Use $u = \sin x$ and $du = \cos x\, dx$.
(b) Replace dx using (a), thus cancelling one power of $\cos x$ by the substitution of du, and be left with an even number of cosine powers.
(c) Use $\cos^2 x = 1 - \sin^2 x$ $(= 1 - u^2)$ to replace the leftover cosines.

3. **Both m and n are odd**:
Use either 1 or 2 (both will work).

4. **Both m and n are even**:
Use $\cos^2 x = \frac{1}{2}(1 + \cos(2x))$ and/or $\sin^2 x = \frac{1}{2}(1 - \cos(2x))$ to reduce to a form that can be integrated.

NOTE: As m and n get large, multiple steps will be needed.

Example 7.12: Odd Power of Cosine and Even Power of Sine

Evaluate $\int \sin^6 x \cos^5 x\, dx$.

Solution. We will show this solution in two ways. First, we show the solution by rewriting the function to make the substitution more clear. Then, we show the solution using the above method.

Solution 1: Rewrite the function as follows.

$$\begin{aligned}
\int \sin^6 x \cos^5 x\, dx &= \int \sin^6 x \cos^4 x \cos x\, dx \\
&= \int \sin^6 x (\cos^2 x)^2 \cos x\, dx \\
&= \int \sin^6 x (1 - \sin^2 x)^2 \cos x\, dx
\end{aligned}$$

Then use the substitution $u = \sin x$ and $du = \cos x\, dx$, that is, $\boxed{dx} = \boxed{\dfrac{du}{\cos x}}$.

$$\begin{aligned}
\int \sin^6 x (1 - \sin^2 x)^2 \cos x\, dx &= \int u^6 (1 - u^2)^2 du \\
&= \int u^6 - 2u^8 + u^{10}\, du
\end{aligned}$$

$$= \frac{u^7}{7} - \frac{2u^9}{9} + \frac{u^{11}}{11} + C$$
$$= \frac{\sin^7 x}{7} - \frac{2\sin^9 x}{9} + \frac{\sin^{11} x}{11} + C$$

Solution 2: Since the power of cosine is odd, we use the substitution $u = \sin x$ and $du = \cos x\, dx$, that is, $\boxed{dx} = \boxed{\dfrac{du}{\cos x}}$.

Then $\int \sin^6 x \cos^5 x\, dx$ is equal to:

$$\begin{aligned}
&= \int u^6 \cos^5 x \boxed{\frac{du}{\cos x}} && \text{Using the substitution}\\
&= \int u^6 \left(\cos^2 x\right)^2 du && \text{Canceling a } \cos x \text{ and rewriting } \cos^4 x\\
&= \int u^6 (1-\sin^2 x)^2\, du && \text{Using trig identity } \cos^2 x = 1-\sin^2 x\\
&= \int u^6 (1-u^2)^2\, du && \text{Writing integral in terms of } u\text{'s}\\
&= \int u^6 - 2u^8 + u^{10}\, du && \text{Expand and collect like terms}\\
&= \frac{u^7}{7} - \frac{2u^9}{9} + \frac{u^{11}}{11} + C && \text{Integrating}\\
&= \frac{\sin^7 x}{7} - \frac{2\sin^9 x}{9} + \frac{\sin^{11} x}{11} + C && \text{Replacing } u \text{ back in terms of } x
\end{aligned}$$

♣

Example 7.13: Odd Power of Cosine

Evaluate $\int \cos^3 x\, dx$.

Solution. Since the power of cosine is odd, we use the substitution $u = \sin x$ and $du = \cos x\, dx$. This may seem strange at first since we don't have $\sin x$ in the question, but it does work!

$$\begin{aligned}
\int \cos^3 x \boxed{dx} &= \int \cos^3 x \boxed{\frac{du}{\cos x}} && \text{Using the substitution}\\
&= \int \cos^2 x\, du && \text{Canceling a } \cos x\\
&= \int (1-\sin^2 x)\, du && \text{Using trig identity } \cos^2 x = 1-\sin^2 x\\
&= \int (1-u^2)\, du && \text{Writing integral in terms of } u\text{'s}\\
&= u - \frac{u^3}{3} + C && \text{Integrating}\\
&= \sin x - \frac{\sin^3 x}{3} + C && \text{Replacing } u \text{ back in terms of } x
\end{aligned}$$

♣

Example 7.14: Product of Even Powers of Sine and Cosine

Evaluate $\int \sin^2 x \cos^2 x\,dx$.

Solution. Use the formulas $\sin^2 x = (1-\cos(2x))/2$ and $\cos^2 x = (1+\cos(2x))/2$ to get:

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

We then have

$$\begin{aligned}
\int \sin^2 x \cos^2 x\,dx &= \int \frac{1-\cos(2x)}{2} \cdot \frac{1+\cos(2x)}{2}\,dx \\
&= \frac{1}{4}\int 1-\cos^2 2x\,dx \\
&= \frac{1}{4}\left(x - \int \cos^2 2x\,dx\right) \\
&= \frac{1}{4}\left(x - \frac{1}{2}\int 1+\cos 4x\,dx\right) \\
&= \frac{1}{4}\left(x - \frac{1}{2}\left(x + \frac{\sin 4x}{4}\right)\right) \\
&= \frac{1}{4}\left(x - \frac{x}{2} - \frac{\sin 4x}{8}\right) + C
\end{aligned}$$

♣

Example 7.15: Even Power of Sine

Evaluate $\int \sin^6 x\,dx$.

Solution. Use $\sin^2 x = (1-\cos(2x))/2$ to rewrite the function:

$$\begin{aligned}
\int \sin^6 x\,dx &= \int (\sin^2 x)^3\,dx \\
&= \int \frac{(1-\cos 2x)^3}{8}\,dx \\
&= \frac{1}{8}\int 1 - 3\cos 2x + 3\cos^2 2x - \cos^3 2x\,dx.
\end{aligned}$$

Now we have four integrals to evaluate. Ignoring the constant for now:

$$\int 1\,dx = x$$

and

$$\int -3\cos 2x\,dx = -\frac{3}{2}\sin 2x$$

are easy. The $\cos^3 2x$ integral is like the previous example:

$$\begin{aligned}
\int -\cos^3 2x\,dx &= \int -\cos 2x\cos^2 2x\,dx \\
&= \int -\cos 2x(1-\sin^2 2x)\,dx \\
&= \int -\frac{1}{2}(1-u^2)\,du \\
&= -\frac{1}{2}\left(u-\frac{u^3}{3}\right) \\
&= -\frac{1}{2}\left(\sin 2x-\frac{\sin^3 2x}{3}\right).
\end{aligned}$$

And finally we use another trigonometric identity, $\cos^2 x = (1+\cos(2x))/2$:

$$\int 3\cos^2 2x\,dx = 3\int \frac{1+\cos 4x}{2}\,dx = \frac{3}{2}\left(x+\frac{\sin 4x}{4}\right).$$

So at long last we get

$$\int \sin^6 x\,dx = \frac{x}{8}-\frac{3}{16}\sin 2x-\frac{1}{16}\left(\sin 2x-\frac{\sin^3 2x}{3}\right)+\frac{3}{16}\left(x+\frac{\sin 4x}{4}\right)+C.$$

♣

Next, we turn our attention to products of secant and tangent. Some we already know how to do.

$$\int \sec^2 x\,dx = \tan x + C \qquad\qquad \int \sec x\tan x\,dx = \sec x + C$$

We can also integrate $\tan x$ quite easily.

Example 7.16: Integrating Tangent

Evaluate $\int \tan x\,dx$.

Solution. Note that $\tan x = \dfrac{\sin x}{\cos x}$ and let $u = \cos x$, so that $du = -\sin x\,dx$.

$$\begin{aligned}
\int \tan x\,dx &= \int \frac{\sin x}{\cos x}\boxed{dx} && \text{Rewriting } \tan x \\
&= \int \frac{\sin x}{u}\boxed{\frac{du}{-\sin x}} && \text{Using the substitution} \\
&= -\int \frac{1}{u}\,du && \text{Cancelling and pulling the } -1 \text{ out} \\
&= -\ln|u|+C && \text{Using formula } \int \frac{1}{u}\,dx = \ln|u|+C \\
&= -\ln|\cos x|+C && \text{Replacing } u \text{ back in terms of } x \\
&= \ln|\sec x|+C && \text{Using log properties and } \sec x = 1/\cos x
\end{aligned}$$

♣

Let's take a moment to realize this result! A common mistake is to believe that $\int \tan x\,dx$ is $\sec^2(x)+C$ – this is *not* true.

Example 7.17: Integrating Tangent Squared

Evaluate $\displaystyle\int \tan^2 x\,dx$.

Solution. Note that $\tan^2 x = \sec^2 x - 1$.

$$\begin{aligned} \int \tan^2 x\,dx &= \int \sec^2 x - 1\,dx && \text{Rewriting } \tan x \\ &= \tan x - x + C && \text{Since } \int \sec^2 x\,dx = \tan x + C \end{aligned}$$

♣

In problems with tangent and secant, two integrals come up frequently:

$$\int \sec^3 x\,dx \qquad \text{and} \qquad \int \sec x\,dx.$$

Both have relatively nice expressions but they are a bit tricky to discover.

First we do $\displaystyle\int \sec x\,dx$, which we will need to compute $\displaystyle\int \sec^3 x\,dx$.

Example 7.18: Integral of Secant

Evaluate $\displaystyle\int \sec x\,dx$.

Solution.

$$\begin{aligned} \int \sec x\,dx &= \int \sec x \frac{\sec x + \tan x}{\sec x + \tan x}\,dx \\ &= \int \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x}\,dx. \end{aligned}$$

Now let $u = \sec x + \tan x$, $du = \sec x \tan x + \sec^2 x\,dx$, exactly the numerator of the function we are integrating. Thus

$$\begin{aligned} \int \sec x\,dx &= \int \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x}\,dx \\ &= \int \frac{1}{u}\,du = \ln|u| + C \\ &= \ln|\sec x + \tan x| + C. \end{aligned}$$

♣

Now we compute the integral $\int \sec^3 x\,dx$.

Example 7.19: Integral of Secant Cubed

Evaluate $\int \sec^3 x\,dx$.

Solution.

$$\begin{aligned}
\sec^3 x &= \frac{\sec^3 x}{2}+\frac{\sec^3 x}{2}=\frac{\sec^3 x}{2}+\frac{(\tan^2 x+1)\sec x}{2}\\
&= \frac{\sec^3 x}{2}+\frac{\sec x\tan^2 x}{2}+\frac{\sec x}{2}\\
&= \frac{\sec^3 x+\sec x\tan^2 x}{2}+\frac{\sec x}{2}.
\end{aligned}$$

We already know how to integrate $\sec x$, so we just need the first quotient. This is “simply” a matter of recognizing the product rule in action:

$$\int \sec^3 x+\sec x\tan^2 x\,dx=\sec x\tan x.$$

So putting these together we get

$$\int \sec^3 x\,dx=\frac{\sec x\tan x}{2}+\frac{\ln|\sec x+\tan x|}{2}+C,$$

♣

For products of secant and tangent it is best to use the following guidelines.

Products of Secant and Tangent

When evaluating $\int \sec^m x \tan^n x\, dx$:

1. **The power of secant is even (m even):**
   (a) Use $u = \tan x$ and $du = \sec^2 x\, dx$.
   (b) Cancel $\sec^2 x$ by the substitution of dx, and be left with an even number of secants.
   (c) Use $\sec^2 x = 1 + \tan^2 x$ $(= 1 + u^2)$ to replace the leftover secants.

2. **The power of tangent is odd (n odd):**
   (a) Use $u = \sec x$ and $du = \sec x \tan x\, dx$.
   (b) Cancel one $\sec x$ and one $\tan x$ by the substitution of dx.
   The number of remaining tangents is even.
   (c) Use $\tan^2 x = \sec^2 x - 1$ $(= u^2 - 1)$ to replace the leftover tangents.

3. **m is even or n is odd**:
   Use either 1 or 2 (both will work).

4. **The power of secant is odd and the power of tangent is even**:
   No guidelines. Remember that $\int \sec x\, dx$ and $\int \sec^3 x\, dx$ can usually be looked up.

Example 7.20: Even Power of Secant

Evaluate $\int \sec^6 x \tan^6 x\, dx$.

Solution. Since the power of secant is even, we ue $u = \tan x$, so that $du = \sec^2 x\, dx$.

$$\begin{aligned}
\int \sec^6 x \tan^6 x\, dx &= \int \sec^6 x\, (u^6) \boxed{\frac{du}{\sec^2 x}} && \text{Using the substitution} \\
&= \int \sec^4 x (u^6)\, du && \text{Cancelling a } \sec^2 x \\
&= \int (\sec^2 x)^2 (u^6)\, du && \text{Rewriting } \sec^4 x \\
&= \int (1 + \tan^2 x)^2 (u^6)\, du && \text{Using } \sec^2 x = 1 + \tan^2 x \\
&= \int (1 + u^2)^2 (u^6)\, du && \text{Using the substitution}
\end{aligned}$$

To integrate this product the easiest method is expand it into a polynomial and integrate term-by-term.

$$\begin{aligned} \int \sec^6 x \tan^6 x\,dx &= \int (u^6 + 2u^8 + u^{10})\,du && \text{Expanding} \\ &= \frac{u^7}{7} + \frac{2u^9}{9} + \frac{u^{11}}{11} + C && \text{Integrating} \\ &= \frac{\tan^7 x}{7} + \frac{2\tan^9 x}{9} + \frac{\tan^{11} x}{11} + C && \text{Rewriting in terms of } x \end{aligned}$$

♣

Example 7.21: Odd Power of Tangent

Evaluate $\int \sec^5 x \tan x\,dx$.

Solution. Since the power of tangent is odd, we use $u = \sec x$, so that $du = \sec x \tan x\,dx$. Then we have:

$$\begin{aligned} \int \sec^5 x \tan x\,dx &= \int \sec^5 x \tan x \boxed{\frac{du}{\sec x \tan x}} && \text{Substituting } dx \text{ first} \\ &= \int \sec^4 x\,du && \text{Cancelling} \\ &= \int u^4\,du && \text{Using the substitution} \\ &= \frac{u^5}{5} + C && \text{Integrating} \\ &= \frac{\sec^5 x}{5} + C && \text{Rewriting in terms of } x \end{aligned}$$

♣

Example 7.22: Odd Power of Secant and Even Power of Tangent

Evaluate $\int \sec x \tan^2 x\,dx$.

Solution. The guidelines don't help us in this scenario. However, since $\tan^2 x = \sec^2 x - 1$, we have

$$\begin{aligned} \int \sec x \tan^2 x\,dx &= \int \sec x(\sec^2 x - 1)\,dx \\ &= \int (\sec^3 x - \sec x)\,dx \\ &= \frac{1}{2}(\sec x \tan x + \ln|\sec x + \tan x|) - \ln|\sec x + \tan x| + C \\ &= \frac{1}{2}\sec x \tan x + \frac{1}{2}\ln|\sec x + \tan x| - \ln|\sec x + \tan x| + C \\ &= \frac{1}{2}\sec x \tan x - \frac{1}{2}\ln|\sec x + \tan x| + C \end{aligned}$$

♣

Exercises for 7.2

Find the antiderivatives.

Exercise 7.2.1 $\int \sin^2 x\,dx$

Exercise 7.2.2 $\int \sin^3 x\,dx$

Exercise 7.2.3 $\int \sin^4 x\,dx$

Exercise 7.2.4 $\int \cos^2 x \sin^3 x\,dx$

Exercise 7.2.5 $\int \cos^3 x\,dx$

Exercise 7.2.6 $\int \cos^3 x \sin^2 x\,dx$

Exercise 7.2.7 $\int \sin x(\cos x)^{3/2}\,dx$

Exercise 7.2.8 $\int \sec^2 x \csc^2 x\,dx$

Exercise 7.2.9 $\int \tan^3 x \sec x\,dx$

Exercise 7.2.10 $\int \left(\frac{1}{\csc x} + \frac{1}{\sec x}\right) dx$

Exercise 7.2.11 $\int \frac{\cos^2 x + \cos x + 1}{\cos^3 x}\,dx$

Exercise 7.2.12 $\int x \sec^2(x^2)\tan^4(x^2)\,dx$

7.3 Trigonometric Substitutions

So far we have seen that it sometimes helps to replace a subexpression of a function by a single variable. Occasionally it can help to replace the original variable by something more complicated. This seems like a "reverse" substitution, but it is really no different in principle than ordinary substitution.

Example 7.23: Sine Subsitution

Evaluate $\int \sqrt{1-x^2}\,dx$.

Solution. Let $x = \sin u$ so $dx = \cos u\,du$. Then

$$\int \sqrt{1-x^2}\,dx = \int \sqrt{1-\sin^2 u}\cos u\,du = \int \sqrt{\cos^2 u}\cos u\,du.$$

We would like to replace $\sqrt{\cos^2 u}$ by $\cos u$, but this is valid only if $\cos u$ is positive, since $\sqrt{\cos^2 u}$ is positive. Consider again the substitution $x = \sin u$. We could just as well think of this as $u = \arcsin x$. If we do, then by the definition of the arcsine, $-\pi/2 \le u \le \pi/2$, so $\cos u \ge 0$. Then we continue:

$$\begin{aligned}
\int \sqrt{\cos^2 u}\cos u\,du &= \int \cos^2 u\,du \\
&= \int \frac{1+\cos 2u}{2}\,du = \frac{u}{2} + \frac{\sin 2u}{4} + C \\
&= \frac{\arcsin x}{2} + \frac{\sin(2\arcsin x)}{4} + C.
\end{aligned}$$

This is a perfectly good answer, though the term $\sin(2\arcsin x)$ is a bit unpleasant. It is possible to simplify this. Using the identity $\sin 2x = 2\sin x\cos x$, we can write $\sin 2u = 2\sin u\cos u = 2\sin(\arcsin x)\sqrt{1-\sin^2 u} = 2x\sqrt{1-\sin^2(\arcsin x)} = 2x\sqrt{1-x^2}$. Then the full antiderivative is

$$\frac{\arcsin x}{2} + \frac{2x\sqrt{1-x^2}}{4} = \frac{\arcsin x}{2} + \frac{x\sqrt{1-x^2}}{2} + C.$$

♣

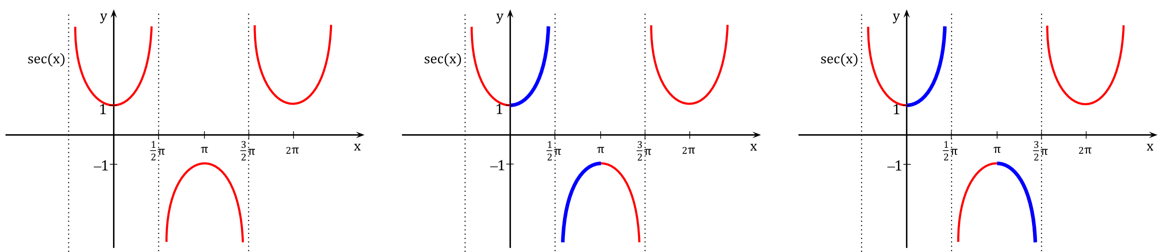
This type of substitution is usually indicated when the function you wish to integrate contains a polynomial expression that might allow you to use the fundamental identity $\sin^2 x + \cos^2 x = 1$ in one of three forms:

$$\cos^2 x = 1 - \sin^2 x \qquad \sec^2 x = 1 + \tan^2 x \qquad \tan^2 x = \sec^2 x - 1.$$

If your function contains $1-x^2$, as in the example above, try $x = \sin u$; if it contains $1+x^2$ try $x = \tan u$; and if it contains x^2-1, try $x = \sec u$. Sometimes you will need to try something a bit different to handle constants other than one which we will describe below. First we discuss inverse substitutions.

In a **traditional** substitution we let $u = u(x)$, i.e., our new variable is defined in terms of x. In an **inverse** substitution we let $x = g(u)$, i.e., we assume x can be written in terms of u. We cannot do this arbitrarily since we do **NOT** get to "choose" x. For example, an inverse substitution of $x = 1$ will give an obviously wrong answer. However, when $x = g(u)$ is an invertible function, then we are really doing a u-substitution with $u = g^{-1}(x)$. Now the substitution rule applies.

Sometimes with inverse substitutions involving trig functions we use θ instead of u. Thus, we would take $x = \sin\theta$ instead of $x = \sin u$. However, as we discussed above, we would like our inverse substitution $x = g(u)$ to be a one-to-one function, and $x = \sin u$ is not one-to-one. We can overcome this issue by using the restricted trigonometric functions. The three common trigonometric substitutions are the restricted sine, restricted tangent and restricted secant. Thus, for sine we use the domain $[-\pi/2,\ \pi/2]$ and for tangent we use $(-\pi/2,\ \pi/2)$. Depending on the convention chosen, the restricted secant function is usually defined in one of two ways.



One convention is to restrict secant to the region $[0,\ \pi/2) \cup (\pi/2, \pi]$ as shown in the middle graph. The other convention is to use $[0,\ \pi/2) \cup [\pi,\ 3\pi/2)$ as shown in the right graph. Both choices give a one-to-one restricted secant function and no universal convention has been adopted. To make the analysis in this section less cumbersome, we will use the domain $[0,\ \pi/2) \cup [\pi,\ 3\pi/2)$ for the restricted secant function. Then $\sec^{-1} x$ is defined to be the inverse of this restricted secant function.

Typically trigonometric substitions are used for problems that involve radical expressions. The table below outlines when each substitution is typically used along with their intervals of validity.

Expression	**Substitution**	**Validity**
$\sqrt{a^2 - x^2}$	$x = a\sin\theta$	$\theta \in [-\pi/2,\ \pi/2]$
$\sqrt{a^2 + x^2}$ or $a^2 + x^2$	$x = a\tan\theta$	$\theta \in (-\pi/2,\ \pi/2)$
$\sqrt{x^2 - a^2}$	$x = a\sec\theta$	$\theta \in [0,\ \pi/2) \cup [\pi,\ 3\pi/2)$

All three substitutions are one-to-one on the listed intervals. When dealing with radicals we often end up with absolute values since

$$\sqrt{z^2} = |z|.$$

For each of the three trigonometric substitions above we will verify that we can ignore the absolute value in each case when encountering a radical.

For $x = a\sin\theta$, the expression $\sqrt{a^2 - x^2}$ becomes

$$\sqrt{a^2 - x^2} = \sqrt{a^2 - a^2\sin^2\theta} = \sqrt{a^2(1 - \sin^2\theta)} = a\sqrt{\cos^2\theta} = a|\cos\theta| = a\cos\theta$$

This is because $\cos\theta \geq 0$ when $\theta \in [-\pi/2,\ \pi/2]$. For $x = a\tan\theta$, the expression $\sqrt{a^2+x^2}$ becomes

$$\sqrt{a^2+x^2} = \sqrt{a^2+a^2\tan^2\theta} = \sqrt{a^2(1+\tan^2\theta)} = a\sqrt{\sec^2\theta} = a|\sec\theta| = a\sec\theta$$

This is because $\sec\theta > 0$ when $\theta \in (-\pi/2,\ \pi/2)$.

Finally, for $x = a\sec\theta$, the expression $\sqrt{x^2-a^2}$ becomes

$$\sqrt{x^2-a^2} = \sqrt{a^2\sec^2\theta - a^2} = \sqrt{a^2(\sec^2\theta-1)} = a\sqrt{\tan^2\theta} = a|\tan\theta| = a\tan\theta$$

This is because $\tan\theta \geq 0$ when $\theta \in [0,\ \pi/2)\cup[\pi,\ 3\pi/2)$.

Thus, when using an appropriate trigonometric substitution we can usually ignore the absolute value. After integrating, we typically get an answer in terms of θ (or u) and need to convert back to x's. To do so, we use the two guidelines below:

- For trig functions containing θ, use a triangle to convert to x's.
- For θ by itself, use the inverse trig function.

All pieces needed for such a trigonometric substitution can be summarized as follows:

Expression	**Substitution**	**Differential**	**Identity**	**Inverse of Substitution**
$\sqrt{a^2-x^2}$	$x = a\sin\theta$	$dx = a\cos\theta d\theta$	$\sqrt{a^2-x^2} = a\cos\theta$	$\theta = \sin^{-1}\left(\frac{x}{a}\right)$
$\sqrt{a^2+x^2}$ or a^2+x^2	$x = a\tan\theta$	$dx = a\sec^2\theta d\theta$	$\sqrt{a^2+x^2} = a\sec\theta$	$\theta = \tan^{-1}\left(\frac{x}{a}\right)$
$\sqrt{x^2-a^2}$	$x = a\sec\theta$	$dx = a\sec\theta\tan\theta d\theta$	$\sqrt{x^2-a^2} = a\tan\theta$	$\theta = \sec^{-1}\left(\frac{x}{a}\right)$

To emphasize the technique, we redo the computation for $\displaystyle\int \sqrt{1-x^2}\,dx$.

Example 7.24: Sine Subsitution

Evaluate $\displaystyle\int \sqrt{1-x^2}\,dx$.

Solution. Since $\sqrt{1-x^2}$ appears in the integrand we try the trigonometric substitution $x = \sin\theta$. (Here we are using the restricted sine function with $\theta \in [-\pi/2,\ \pi/2]$ but typically omit this detail when writing out the solution.) Then $\boxed{dx} = \boxed{\cos\theta\, d\theta}$.

$$\begin{aligned}
\int \sqrt{1-x^2}\,\boxed{dx} &= \int \sqrt{1-\sin^2\theta}\,\boxed{\cos\theta\, d\theta} && \text{Using our (inverse) substitution}\\
&= \int \sqrt{\cos^2\theta}\cos\theta\, d\theta && \text{Since } \sin^2\theta+\cos^2\theta=1\\
&= \int |\cos\theta|\cdot\cos\theta\, d\theta && \text{Since } \sqrt{\cos^2\theta}=|\cos\theta|\\
&= \int \cos^2\theta\, d\theta && \text{Since for } \theta\in[-\tfrac{\pi}{2},\tfrac{\pi}{2}] \text{ we have } \cos\theta\geq 0.
\end{aligned}$$

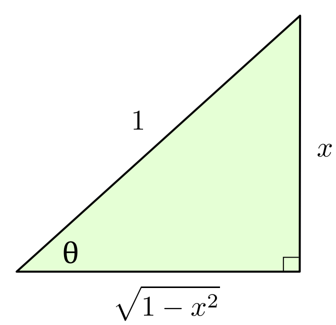
Often we omit the step containing the absolute value by our discussion above. Now, to integrate a power of cosine we use the guidelines for products of sine and cosine and make use of the identity

$$\cos^2\theta = \frac{1}{2}(1+\cos(2\theta)).$$

Our integral then becomes

$$\int \sqrt{1-x^2}\,dx = \frac{1}{2}\int (1+\cos(2\theta))\,d\theta = \frac{\theta}{2}+\frac{\sin(2\theta)}{4}+C$$

To write the answer back in terms of x we use a right triangle. Since $\sin\theta = x/1$ we have the triangle:



The triangle gives $\sin\theta$, $\cos\theta$, $\tan\theta$, but have a $\sin(2\theta)$. Thus, we use an identity to write

$$\sin(2\theta) = 2\sin\theta\cos\theta = 2\left(\frac{x}{1}\right)\left(\frac{\sqrt{1-x^2}}{1}\right)$$

For θ by itself we use $\theta = \sin^{-1}x$. Thus, the integral is

$$\int \sqrt{1-x^2}\,dx = \frac{\sin^{-1}x}{2}+\frac{x\sqrt{1-x^2}}{2}+C$$

♣

Example 7.25: Secant Subsitution

Evaluate $\displaystyle\int \frac{\sqrt{25x^2-4}}{x}\,dx.$

Solution. We do not have $\sqrt{x^2-a^2}$ because of the 25, but if we factor 25 out we get:

$$\int \frac{\sqrt{25(x^2-(4/25))}}{x}\,dx = \int 5\frac{\sqrt{x^2-(4/25)}}{x}\,dx.$$

Now, $a = 2/5$, so let $x = \frac{2}{5}\sec\theta$. Alternatively, we can think of the integral as being:

$$\int \frac{\sqrt{(5x)^2-4}}{x}\,dx$$

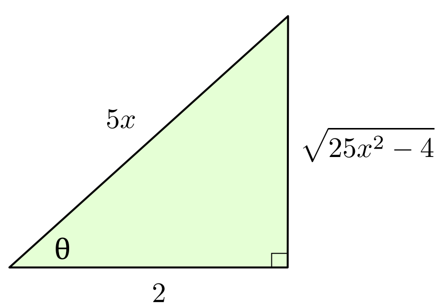
Then we could let $u = 5x$ followed by $u = 2\sec\theta$, etc. Or equivalently, we can avoid a u-substitution by letting $5x = 2\sec\theta$. In either case we are using the trigonometric substitution $x = \frac{2}{5}\sec\theta$, but do use the method that makes the most sense to you! As $x = \frac{2}{5}\sec\theta$ we have $\boxed{dx} = \boxed{\frac{2}{5}\sec\theta\tan\theta\,d\theta}$.

$$\begin{aligned}
\int \frac{\sqrt{25x^2-4}}{x}\boxed{dx} &= \int \frac{\sqrt{25\frac{4\sec^2\theta}{25}-4}}{\frac{2}{5}\sec\theta}\boxed{\frac{2}{5}\sec\theta\tan\theta\,d\theta} && \text{Using the substitution}\\
&= \int \sqrt{4(\sec^2\theta-1)}\cdot\tan\theta\,d\theta && \text{Cancelling}\\
&= 2\int \sqrt{\tan^2\theta}\cdot\tan\theta\,d\theta && \text{Using } \tan^2\theta+1=\sec^2\theta\\
&= 2\int \tan^2\theta\,d\theta && \text{Simplifying}\\
&= 2\int (\sec^2\theta-1)\,d\theta && \text{Using } \tan^2\theta+1=\sec^2\theta\\
&= 2(\tan\theta-\theta)+C && \text{Since } \int \sec^2\theta\,d\theta = \tan\theta+C
\end{aligned}$$

For $\tan\theta$, we use a right triangle.

$$x = \frac{2}{5}\sec\theta \qquad \to \qquad x = \frac{2}{5}\frac{1}{\cos\theta} \qquad \to \qquad \cos\theta = \frac{2}{5x}$$

Using SOH CAH TOA, the triangle is then



For θ by itself, we use $\theta = \sec^{-1}(5x/2)$. Thus,

$$\int \frac{\sqrt{25x^2-4}}{x}\,dx = 2\left(\frac{\sqrt{25x^2-4}}{2} - \sec^{-1}\left(\frac{5x}{2}\right)\right)+C$$

♣

In the context of the previous example, some resources give alternate guidelines when choosing a trigonometric substitution.

$$\sqrt{a^2 - b^2x^2} \quad \rightarrow \quad x = \frac{a}{b}\sin\theta$$

$$\sqrt{b^2x^2 + a^2} \text{ or } (b^2x^2 + a^2) \quad \rightarrow \quad x = \frac{a}{b}\tan\theta$$

$$\sqrt{b^2x^2 - a^2} \quad \rightarrow \quad x = \frac{a}{b}\sec\theta$$

We next look at a tangent substitution.

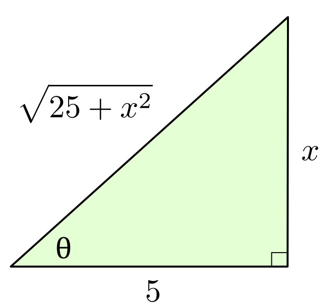
Example 7.26: Tangent Substitution

Evaluate $\int \frac{1}{\sqrt{25+x^2}}\,dx$.

Solution. Let $x = 5\tan\theta$ so that $\boxed{dx} = \boxed{5\sec^2\theta\,d\theta}$.

$$\begin{aligned}
\int \frac{1}{\sqrt{25+x^2}}\boxed{dx} &= \int \frac{1}{\sqrt{25+25\tan^2\theta}}\boxed{5\sec^2\theta\,d\theta} && \text{Using our substitution} \\
&= \int \frac{1}{\sqrt{25(1+\tan^2\theta)}} \cdot 5\sec^2\theta\,d\theta && \text{Factor out 25} \\
&= \int \frac{1}{5\sqrt{sec^2\theta}} \cdot 5\sec^2\theta\,d\theta && \text{Using } \tan^2\theta + 1 = \sec^2\theta \\
&= \int \sec\theta\,d\theta && \text{Simplifying} \\
&= \ln|\sec\theta + \tan\theta| + C && \text{By } \int \sec\theta\,dx = \ln|\sec\theta + \tan\theta| + C
\end{aligned}$$

Since $\tan\theta = x/5$, we draw a triangle:



Then

$$\sec\theta = \frac{1}{\cos\theta} = \frac{\sqrt{25+x^2}}{5}.$$

Therefore, the integral is

$$\int \frac{1}{\sqrt{25+x^2}}\,dx = \ln\left|\frac{\sqrt{25+x^2}}{5} + \frac{x}{5}\right| + C$$

♣

In the next example, we will use the technique of completing the square in order to rewrite the integrand.

Example 7.27: Completing the Square

Evaluate $\displaystyle\int \frac{x}{\sqrt{3-2x-x^2}}\,dx.$

Solution. First, complete the square to write

$$3-2x-x^2 = 4-(x+1)^2$$

Now, we may let $u = x+1$ so that $du = dx$ (note that $x = u-1$) to get:

$$\int \frac{x}{\sqrt{4-(x+1)^2}}\,dx = \int \frac{u-1}{\sqrt{4-u^2}}\,du$$

Let $u = 2\sin\theta$ giving $du = 2\cos\theta\, d\theta$:

$$\int \frac{u-1}{\sqrt{4-u^2}}\,du = \int \frac{2\sin\theta - 1}{2\cos\theta}\cdot 2\cos\theta\, d\theta = \int (2\sin\theta - 1)\, d\theta$$

Integrating and using a triangle we get:

$$\begin{aligned}
\int \frac{x}{\sqrt{3-2x-x^2}} &= -2\cos\theta - \theta + C \\
&= -\sqrt{4-u^2} - \sin^{-1}\left(\frac{u}{2}\right) + C \\
&= -\sqrt{3-2x-x^2} - \sin^{-1}\left(\frac{x+1}{2}\right) + C
\end{aligned}$$

Note that in this problem we could have skipped the u-substitution if instead we let $x+1 = 2\sin\theta$. (For the triangle we would then use $\sin\theta = \dfrac{x+1}{2}$.) ♣

Exercises for 7.3

Exercise 7.3.1 $\displaystyle\int \sqrt{x^2-1}\,dx$

Exercise 7.3.2 $\displaystyle\int \sqrt{9+4x^2}\,dx$

Exercise 7.3.3 $\displaystyle\int x\sqrt{1-x^2}\,dx$

Exercise 7.3.4 $\int x^2\sqrt{1-x^2}\,dx$

Exercise 7.3.5 $\int \frac{1}{\sqrt{1+x^2}}\,dx$

Exercise 7.3.6 $\int \sqrt{x^2+2x}\,dx$

Exercise 7.3.7 $\int \frac{1}{x^2(1+x^2)}\,dx$

Exercise 7.3.8 $\int \frac{x^2}{\sqrt{4-x^2}}\,dx$

Exercise 7.3.9 $\int \frac{\sqrt{x}}{\sqrt{1-x}}\,dx$

Exercise 7.3.10 $\int \frac{x^3}{\sqrt{4x^2-1}}\,dx$

7.4 Integration by Parts

We have already seen that recognizing the product rule can be useful, when we noticed that

$$\int \sec^3 u + \sec u \tan^2 u\,du = \sec u \tan u.$$

As with substitution, we do not have to rely on insight or cleverness to discover such antiderivatives; there is a technique that will often help to uncover the product rule.

Start with the product rule:

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

We can rewrite this as

$$f(x)g(x) = \int f'(x)g(x)\,dx + \int f(x)g'(x)\,dx,$$

and then

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx.$$

This may not seem particularly useful at first glance, but it turns out that in many cases we have an integral of the form

$$\int f(x)g'(x)\,dx$$

but that

$$\int f'(x)g(x)\,dx$$

is easier. This technique for turning one integral into another is called **integration by parts**, and is usually written in more compact form. If we let $u = f(x)$ and $v = g(x)$ then $du = f'(x)\,dx$ and $dv = g'(x)\,dx$ and

$$\int u\,dv = uv - \int v\,du.$$

To use this technique we need to identify likely candidates for $u = f(x)$ and $dv = g'(x)\,dx$.

Example 7.28: Product of a Linear Function and Logarithm

Evaluate $\int x \ln x\,dx$.

Solution. Let $u = \ln x$ so $du = 1/x\,dx$. Then we must let $dv = x\,dx$ so $v = x^2/2$ and

$$\int x \ln x\,dx = \frac{x^2 \ln x}{2} - \int \frac{x^2}{2}\frac{1}{x}\,dx = \frac{x^2 \ln x}{2} - \int \frac{x}{2}\,dx = \frac{x^2 \ln x}{2} - \frac{x^2}{4} + C.$$

♣

Example 7.29: Product of a Linear Function and Trigonometric Function

Evaluate $\int x \sin x\,dx$.

Solution. Let $u = x$ so $du = dx$. Then we must let $dv = \sin x\,dx$ so $v = -\cos x$ and

$$\int x \sin x\,dx = -x\cos x - \int -\cos x\,dx = -x\cos x + \int \cos x\,dx = -x\cos x + \sin x + C.$$

♣

Example 7.30: Secant Cubed (again)

Evaluate $\int \sec^3 x\,dx$.

Solution. Of course we already know the answer to this, but we needed to be clever to discover it. Here we'll use the new technique to discover the antiderivative. Let $u = \sec x$ and $dv = \sec^2 x\,dx$. Then $du = \sec x \tan x$ and $v = \tan x$ and

$$\begin{aligned}
\int \sec^3 x\,dx &= \sec x \tan x - \int \tan^2 x \sec x\,dx \\
&= \sec x \tan x - \int (\sec^2 x - 1)\sec x\,dx \\
&= \sec x \tan x - \int \sec^3 x\,dx + \int \sec x\,dx.
\end{aligned}$$

At first this looks useless—we're right back to $\int \sec^3 x\,dx$. But looking more closely:

$$\begin{aligned}
\int \sec^3 x\,dx &= \sec x \tan x - \int \sec^3 x\,dx + \int \sec x\,dx \\
\int \sec^3 x\,dx + \int \sec^3 x\,dx &= \sec x \tan x + \int \sec x\,dx \\
2\int \sec^3 x\,dx &= \sec x \tan x + \int \sec x\,dx \\
\int \sec^3 x\,dx &= \frac{\sec x \tan x}{2} + \frac{1}{2}\int \sec x\,dx \\
&= \frac{\sec x \tan x}{2} + \frac{\ln|\sec x + \tan x|}{2} + C.
\end{aligned}$$

♣

Example 7.31: Product of a Polynomial and Trigonometric Function

Evaluate $\int x^2 \sin x\,dx$.

Solution. Let $u = x^2$, $dv = \sin x\,dx$; then $du = 2x\,dx$ and $v = -\cos x$. Now

$$\int x^2 \sin x\,dx = -x^2 \cos x + \int 2x \cos x\,dx.$$

This is better than the original integral, but we need to do integration by parts again. Let $u = 2x$, $dv = \cos x\,dx$; then $du = 2$ and $v = \sin x$, and

$$\begin{aligned}
\int x^2 \sin x\,dx &= -x^2 \cos x + \int 2x \cos x\,dx \\
&= -x^2 \cos x + 2x \sin x - \int 2 \sin x\,dx \\
&= -x^2 \cos x + 2x \sin x + 2\cos x + C.
\end{aligned}$$

♣

Such repeated use of integration by parts is fairly common, but it can be a bit tedious to accomplish, and it is easy to make errors, especially sign errors involving the subtraction in the formula. There is a nice tabular method to accomplish the calculation that minimizes the chance for error and speeds up the whole process. We illustrate with the previous example. Here is the table:

sign	u	dv
+	x^2	$\sin x$
-	$2x$	$-\cos x$
+	2	$-\sin x$
-	0	$\cos x$

To form this table, we start with u at the top of the second column and repeatedly compute the derivative; starting with dv at the top of the third column, we repeatedly compute the antiderivative. In the first column, we place a "$-$" in every second row. To form the second table we combine the first and second columns by ignoring the boundary; if you do this by hand, you may simply start with two columns and add a "−" to every second row.

Alternatively, we can use the following table:

u	dv
x^2	$\sin x$
$-2x$	$-\cos x$
2	$-\sin x$
0	$\cos x$

To compute with this second table we begin at the top. Multiply the first entry in column u by the second entry in column dv to get $-x^2\cos x$, and add this to the integral of the product of the second entry in column u and second entry in column dv. This gives:

$$-x^2\cos x+\int 2x\cos x\,dx,$$

or exactly the result of the first application of integration by parts. Since this integral is not yet easy, we return to the table. Now we multiply twice on the diagonal, $(x^2)(-\cos x)$ and $(-2x)(-\sin x)$ and then once straight across, $(2)(-\sin x)$, and combine these as

$$-x^2\cos x+2x\sin x-\int 2\sin x\,dx,$$

giving the same result as the second application of integration by parts. While this integral is easy, we may return yet once more to the table. Now multiply three times on the diagonal to get $(x^2)(-\cos x)$, $(-2x)(-\sin x)$, and $(2)(\cos x)$, and once straight across, $(0)(\cos x)$. We combine these as before to get

$$-x^2\cos x+2x\sin x+2\cos x+\int 0\,dx=-x^2\cos x+2x\sin x+2\cos x+C.$$

Typically we would fill in the table one line at a time, until the "straight across" multiplication gives an easy integral. If we can see that the u column will eventually become zero, we can instead fill in the whole table; computing the products as indicated will then give the entire integral, including the "$+C$", as above.

Exercises for 7.4

Find the antiderivatives.

Exercise 7.4.1 $\int x\cos x\,dx$

Exercise 7.4.2 $\int x^2\cos x\,dx$

Exercise 7.4.3 $\int xe^x\,dx$

Exercise 7.4.4 $\int xe^{x^2}\,dx$

Exercise 7.4.5 $\int \sin^2 x\,dx$

Exercise 7.4.6 $\int \ln x\,dx$

Exercise 7.4.7 $\int x\arctan x\,dx$

Exercise 7.4.8 $\int x^3\sin x\,dx$

Exercise 7.4.9 $\int x^3\cos x\,dx$

Exercise 7.4.10 $\int x\sin^2 x\,dx$

Exercise 7.4.11 $\int x\sin x\cos x\,dx$

Exercise 7.4.12 $\int \arctan(\sqrt{x})\,dx$

Exercise 7.4.13 $\int \sin(\sqrt{x})\,dx$

Exercise 7.4.14 $\int \sec^2 x\csc^2 x\,dx$

7.5 Rational Functions

A **rational function** is a fraction with polynomials in the numerator and denominator. For example,

$$\frac{x^3}{x^2+x-6}, \qquad \frac{1}{(x-3)^2}, \qquad \frac{x^2+1}{x^2-1},$$

are all rational functions of x. There is a general technique called "partial fractions" that, in principle, allows us to integrate any rational function. The algebraic steps in the technique are rather cumbersome if the polynomial in the denominator has degree more than 2, and the technique requires that we factor the denominator, something that is not always possible. However, in practice one does not often run across rational functions with high degree polynomials in the denominator for which one has to find the antiderivative function. So we shall explain how to find the antiderivative of a rational function only when the denominator is a quadratic polynomial ax^2+bx+c.

We should mention a special type of rational function that we already know how to integrate: If the denominator has the form $(ax+b)^n$, the substitution $u = ax+b$ will always work. The denominator becomes u^n, and each x in the numerator is replaced by $(u-b)/a$, and $dx = du/a$. While it may be tedious to complete the integration if the numerator has high degree, it is merely a matter of algebra.

Example 7.32: Substitution and Splitting Up a Fraction

Find $\displaystyle\int \frac{x^3}{(3-2x)^5}\,dx$.

Solution. Using the substitution $u = 3-2x$ we get

$$\begin{aligned}
\int \frac{x^3}{(3-2x)^5}\,dx &= \frac{1}{-2}\int \frac{\left(\frac{u-3}{-2}\right)^3}{u^5}\,du = \frac{1}{16}\int \frac{u^3-9u^2+27u-27}{u^5}\,du \\
&= \frac{1}{16}\int u^{-2}-9u^{-3}+27u^{-4}-27u^{-5}\,du \\
&= \frac{1}{16}\left(\frac{u^{-1}}{-1}-\frac{9u^{-2}}{-2}+\frac{27u^{-3}}{-3}-\frac{27u^{-4}}{-4}\right)+C \\
&= \frac{1}{16}\left(\frac{(3-2x)^{-1}}{-1}-\frac{9(3-2x)^{-2}}{-2}+\frac{27(3-2x)^{-3}}{-3}-\frac{27(3-2x)^{-4}}{-4}\right)+C \\
&= -\frac{1}{16(3-2x)}+\frac{9}{32(3-2x)^2}-\frac{9}{16(3-2x)^3}+\frac{27}{64(3-2x)^4}+C
\end{aligned}$$

♣

We now proceed to the case in which the denominator is a quadratic polynomial. We can always factor out the coefficient of x^2 and put it outside the integral, so we can assume that the denominator has the form x^2+bx+c. There are three possible cases, depending on how the quadratic factors: either $x^2+bx+c = (x-r)(x-s)$, $x^2+bx+c = (x-r)^2$, or it doesn't factor. We can use the quadratic formula to decide which of these we have, and to factor the quadratic if it is possible.

Example 7.33: Factoring a Quadratic

Determine whether x^2+x+1 *factors, and factor it if possible.*

Solution. The quadratic formula tells us that $x^2+x+1=0$ when

$$x = \frac{-1\pm\sqrt{1-4}}{2}.$$

Since there is no square root of -3, this quadratic does not factor. ♣

Example 7.34: Factoring a Quadratic with Real Roots

Determine whether x^2-x-1 *factors, and factor it if possible.*

Solution. The quadratic formula tells us that $x^2 - x - 1 = 0$ when

$$x = \frac{1 \pm \sqrt{1+4}}{2} = \frac{1 \pm \sqrt{5}}{2}.$$

Therefore

$$x^2 - x - 1 = \left(x - \frac{1+\sqrt{5}}{2}\right)\left(x - \frac{1-\sqrt{5}}{2}\right).$$

♣

If $x^2 + bx + c = (x-r)^2$ then we have the special case we have already seen, that can be handled with a substitution. The other two cases require different approaches.

If $x^2 + bx + c = (x-r)(x-s)$, we have an integral of the form

$$\int \frac{p(x)}{(x-r)(x-s)}\,dx$$

where $p(x)$ is a polynomial. The first step is to make sure that $p(x)$ has degree less than 2.

Example 7.35:

Rewrite

$$\int \frac{x^3}{(x-2)(x+3)}\,dx$$

in terms of an integral with a numerator that has degree less than 2.

Solution. To do this we use long division of polynomials to discover that

$$\frac{x^3}{(x-2)(x+3)} = \frac{x^3}{x^2+x-6} = x - 1 + \frac{7x-6}{x^2+x-6} = x - 1 + \frac{7x-6}{(x-2)(x+3)}.$$

See `http://en.wikipedia.org/wiki/Polynomial_long_division` for a review on long division. Then

$$\int \frac{x^3}{(x-2)(x+3)}\,dx = \int x - 1\,dx + \int \frac{7x-6}{(x-2)(x+3)}\,dx.$$

The first integral is easy, so only the second requires some work. ♣

Now consider the following simple algebra of fractions:

$$\frac{A}{x-r} + \frac{B}{x-s} = \frac{A(x-s)+B(x-r)}{(x-r)(x-s)} = \frac{(A+B)x - As - Br}{(x-r)(x-s)}.$$

That is, adding two fractions with constant numerator and denominators $(x-r)$ and $(x-s)$ produces a fraction with denominator $(x-r)(x-s)$ and a polynomial of degree less than 2 for the numerator. We want to reverse this process: Starting with a single fraction, we want to write it as a sum of two simpler fractions. An example should make it clear how to proceed.

Example 7.36: Partial Fraction Decomposition

Evaluate $\displaystyle\int \frac{x^3}{(x-2)(x+3)}\,dx.$

Solution. We start by writing $\dfrac{7x-6}{(x-2)(x+3)}$ as the sum of two fractions. We want to end up with

$$\frac{7x-6}{(x-2)(x+3)} = \frac{A}{x-2} + \frac{B}{x+3}.$$

If we go ahead and add the fractions on the right hand side, we seek a common denominator, and get:

$$\frac{7x-6}{(x-2)(x+3)} = \frac{(A+B)x+3A-2B}{(x-2)(x+3)}.$$

So all we need to do is find A and B so that $7x-6=(A+B)x+3A-2B$, which is to say, we need $7=A+B$ and $-6=3A-2B$. This is a problem you've seen before: Solve a system of two equations in two unknowns. There are many ways to proceed; here's one: If $7=A+B$ then $B=7-A$ and so $-6=3A-2B=3A-2(7-A)=3A-14+2A=5A-14$. This is easy to solve for A: $A=8/5$, and then $B=7-A=7-8/5=27/5$. Thus

$$\int \frac{7x-6}{(x-2)(x+3)}\,dx = \int \frac{8}{5}\frac{1}{x-2} + \frac{27}{5}\frac{1}{x+3}\,dx = \frac{8}{5}\ln|x-2| + \frac{27}{5}\ln|x+3| + C.$$

The answer to the original problem is now

$$\begin{aligned}\int \frac{x^3}{(x-2)(x+3)}\,dx &= \int x-1\,dx + \int \frac{7x-6}{(x-2)(x+3)}\,dx\\ &= \frac{x^2}{2} - x + \frac{8}{5}\ln|x-2| + \frac{27}{5}\ln|x+3| + C.\end{aligned}$$

♣

Now suppose that x^2+bx+c doesn't factor. Again we can use long division to ensure that the numerator has degree less than 2, then we complete the square.

Example 7.37: Denominator Does Not Factor

Evaluate $\displaystyle\int \frac{x+1}{x^2+4x+8}\,dx.$

Solution. The quadratic denominator does not factor. We could complete the square and use a trigonometric substitution, but it is simpler to rearrange the integrand:

$$\int \frac{x+1}{x^2+4x+8}\,dx = \int \frac{x+2}{x^2+4x+8}\,dx - \int \frac{1}{x^2+4x+8}\,dx.$$

The first integral is an easy substitution problem, using $u=x^2+4x+8$:

$$\int \frac{x+2}{x^2+4x+8}\,dx = \frac{1}{2}\int \frac{du}{u} = \frac{1}{2}\ln|x^2+4x+8|.$$

For the second integral we complete the square:

$$x^2+4x+8=(x+2)^2+4=4\left(\left(\frac{x+2}{2}\right)^2+1\right),$$

making the integral

$$\frac{1}{4}\int\frac{1}{\left(\frac{x+2}{2}\right)^2+1}\,dx.$$

Using $u=\dfrac{x+2}{2}$ we get

$$\frac{1}{4}\int\frac{1}{\left(\frac{x+2}{2}\right)^2+1}\,dx=\frac{1}{4}\int\frac{2}{u^2+1}\,dx=\frac{1}{2}\arctan\left(\frac{x+2}{2}\right).$$

The final answer is now

$$\int\frac{x+1}{x^2+4x+8}\,dx=\frac{1}{2}\ln|x^2+4x+8|-\frac{1}{2}\arctan\left(\frac{x+2}{2}\right)+C.$$

♣

Exercises for 7.5

Exercise 7.5.1 $\displaystyle\int\frac{1}{4-x^2}\,dx$

Exercise 7.5.2 $\displaystyle\int\frac{x^4}{4-x^2}\,dx$

Exercise 7.5.3 $\displaystyle\int\frac{1}{x^2+10x+25}\,dx$

Exercise 7.5.4 $\displaystyle\int\frac{x^2}{4-x^2}\,dx$

Exercise 7.5.5 $\displaystyle\int\frac{x^4}{4+x^2}\,dx$

Exercise 7.5.6 $\displaystyle\int\frac{1}{x^2+10x+29}\,dx$

Exercise 7.5.7 $\displaystyle\int\frac{x^3}{4+x^2}\,dx$

Exercise 7.5.8 $\displaystyle\int \frac{1}{x^2+10x+21}\,dx$

Exercise 7.5.9 $\displaystyle\int \frac{1}{2x^2-x-3}\,dx$

Exercise 7.5.10 $\displaystyle\int \frac{1}{x^2+3x}\,dx$

7.6 Numerical Integration

We have now seen some of the most generally useful methods for discovering antiderivatives, and there are others. Unfortunately, some functions have no simple antiderivatives. In such cases, if the value of a definite integral is needed it will have to be approximated. We will see two methods that work reasonably well and yet are fairly simple; in some cases more sophisticated techniques will be needed.

Of course, we already know one way to approximate an integral: If we think of the integral as computing an area, we can add up the areas of some rectangles. While this is quite simple, it is usually the case that a large number of rectangles is needed to get acceptable accuracy. A similar approach is much better. We approximate the area under a curve over a small interval as the area of a trapezoid. In figure 7.1 we see an area under a curve approximated by rectangles and by trapezoids; it is apparent that the trapezoids give a substantially better approximation on each subinterval.

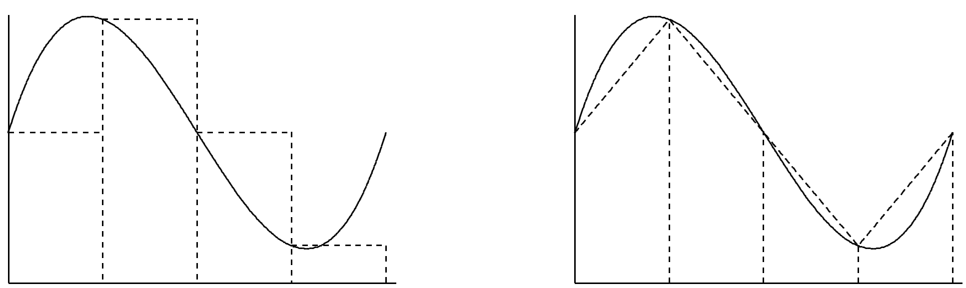


Figure 7.1: Approximating an area with rectangles and with trapezoids.

As with rectangles, we divide the interval into n equal subintervals of length Δx. A typical trapezoid is pictured in figure 7.2; it has area

$$\frac{f(x_i)+f(x_{i+1})}{2}\Delta x.$$

If we add up the areas of all trapezoids we get

$$\begin{aligned}&\frac{f(x_0)+f(x_1)}{2}\Delta x+\frac{f(x_1)+f(x_2)}{2}\Delta x+\cdots+\frac{f(x_{n-1})+f(x_n)}{2}\Delta x\\ &=\left(\frac{f(x_0)}{2}+f(x_1)+f(x_2)+\cdots+f(x_{n-1})+\frac{f(x_n)}{2}\right)\Delta x.\end{aligned}$$

For a modest number of subintervals this is not too difficult to do with a calculator; a computer can easily do many subintervals.

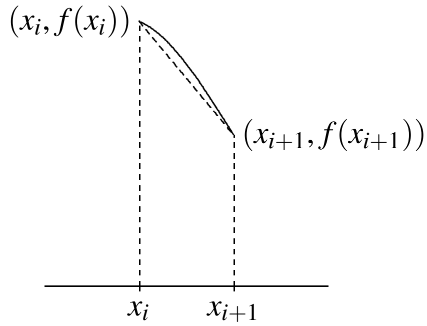


Figure 7.2: A single trapezoid.

In practice, an approximation is useful only if we know how accurate it is; for example, we might need a particular value accurate to three decimal places. When we compute a particular approximation to an integral, the error is the difference between the approximation and the true value of the integral. For any approximation technique, we need an **error bound**, a value that is guaranteed to be larger than the actual error. If A is an approximation and E is the associated error bound, then we know that the true value of the integral is between $A-E$ and $A+E$. In the case of our approximation of the integral, we want $E = E(\Delta x)$ to be a function of Δx that gets small rapidly as Δx gets small. Fortunately, for many functions, there is such an error bound associated with the trapezoid approximation.

Theorem 7.38: Error for Trapezoid Approximation

Suppose f has a second derivative f'' everywhere on the interval $[a,b]$, and $|f''(x)| \le M$ for all x in the interval. With $\Delta x = (b-a)/n$, an error bound for the trapezoid approximation is

$$E(\Delta x) = \frac{b-a}{12} M(\Delta x)^2 = \frac{(b-a)^3}{12n^2} M.$$

Let's see how we can use this.

Example 7.39: Approximate an Integral With Trapezoids

Approximate $\int_0^1 e^{-x^2}\,dx$ to two decimal places.

Solution. The second derivative of $f = e^{-x^2}$ is $(4x^2-2)e^{-x^2}$, and it is not hard to see that on $[0,1]$ $|f''(x)|$ has a maximum value of 2, thus we begin by estimating the number of subintervals we are likely to need. To get two decimal places of accuracy, we will certainly need $E(\Delta x) < 0.005$ or

$$\begin{aligned} \frac{1}{12}(2)\frac{1}{n^2} &< 0.005 \\ \frac{1}{6}(200) &< n^2 \\ 5.77 \approx \sqrt{\frac{100}{3}} &< n \end{aligned}$$

With $n = 6$, the error bound is thus $1/6^3 < 0.0047$. We compute the trapezoid approximation for six intervals:

$$\left(\frac{f(0)}{2} + f(1/6) + f(2/6) + \cdots + f(5/6) + \frac{f(1)}{2}\right)\frac{1}{6} \approx 0.74512.$$

So the true value of the integral is between $0.74512 - 0.0047 = 0.74042$ and $0.74512 + 0.0047 = 0.74982$. Unfortunately, the first rounds to 0.74 and the second rounds to 0.75, so we can't be sure of the correct value in the second decimal place; we need to pick a larger n. As it turns out, we need to go to $n = 12$ to get two bounds that both round to the same value, which turns out to be 0.75. For comparison, using 12 rectangles to approximate the area gives 0.7727, which is considerably less accurate than the approximation using six trapezoids.

In practice it generally pays to start by requiring better than the maximum possible error; for example, we might have initially required $E(\Delta x) < 0.001$, or

$$\begin{aligned} \frac{1}{12}(2)\frac{1}{n^2} &< 0.001 \\ \frac{1}{6}(1000) &< n^2 \\ 12.91 \approx \sqrt{\frac{500}{3}} &< n \end{aligned}$$

Had we immediately tried $n = 13$ this would have given us the desired answer. ♣

The trapezoid approximation works well, especially compared to rectangles, because the tops of the trapezoids form a reasonably good approximation to the curve when Δx is fairly small. We can extend this idea: what if we try to approximate the curve more closely by using something other than a straight line? The obvious candidate is a parabola: If we can approximate a short piece of the curve with a parabola with equation $y = ax^2 + bx + c$, we can easily compute the area under the parabola.

There are an infinite number of parabolas through any two given points, but only one through three given points. If we find a parabola through three consecutive points $(x_i, f(x_i))$, $(x_{i+1}, f(x_{i+1}))$, $(x_{i+2}, f(x_{i+2}))$ on the curve, it should be quite close to the curve over the whole interval $[x_i, x_{i+2}]$, as in Figure 7.3. If we divide the interval $[a,b]$ into an even number of subintervals, we can then approximate the curve by a sequence of parabolas, each covering two of the subintervals. For this to be practical, we would like a simple formula for the area under one parabola, namely, the parabola through $(x_i, f(x_i))$, $(x_{i+1}, f(x_{i+1}))$, and $(x_{i+2}, f(x_{i+2}))$. That is, we should attempt to write down the parabola $y = ax^2 + bx + c$ through these points and then integrate it, and hope that the result is fairly simple. Although the algebra involved is messy, this turns out to be possible. The algebra is well within the capability of a good computer algebra system like Sage, so we will present the result without all of the algebra.

To find the parabola, we solve these three equations for a, b, and c:

$$\begin{aligned} f(x_i) &= a(x_{i+1} - \Delta x)^2 + b(x_{i+1} - \Delta x) + c \\ f(x_{i+1}) &= a(x_{i+1})^2 + b(x_{i+1}) + c \\ f(x_{i+2}) &= a(x_{i+1} + \Delta x)^2 + b(x_{i+1} + \Delta x) + c \end{aligned}$$

Not surprisingly, the solutions turn out to be quite messy. Nevertheless, Sage can easily compute and simplify the integral to get

$$\int_{x_{i+1}-\Delta x}^{x_{i+1}+\Delta x} ax^2 + bx + c\,dx = \frac{\Delta x}{3}(f(x_i) + 4f(x_{i+1}) + f(x_{i+2})).$$

Now the sum of the areas under all parabolas is

$$\frac{\Delta x}{3}(f(x_0)+4f(x_1)+f(x_2)+f(x_2)+4f(x_3)+f(x_4)+\cdots+f(x_{n-2})+4f(x_{n-1})+f(x_n)) =$$
$$\frac{\Delta x}{3}(f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)+2f(x_4)+\cdots+2f(x_{n-2})+4f(x_{n-1})+f(x_n)).$$

This is just slightly more complicated than the formula for trapezoids; we need to remember the alternating 2 and 4 coefficients, and that the interval must be divided into an *even* number of subintervals. This approximation technique is referred to as **Simpson's Rule**.

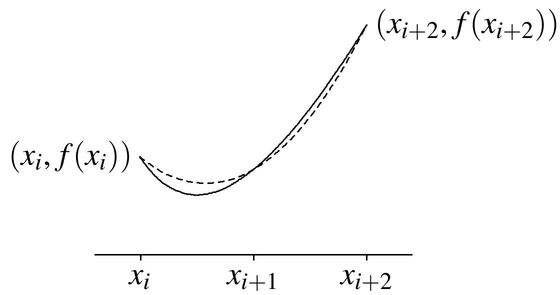


Figure 7.3: A parabola (dashed) approximating a curve (solid).

As with the trapezoid method, this is useful only with an error bound:

Theorem 7.40: Error for Simpson's Approximation

Suppose f has a fourth derivative $f^{(4)}$ everywhere on the interval $[a,b]$, and $|f^{(4)}(x)| \le M$ for all x in the interval. With $\Delta x = (b-a)/n$, an error bound for Simpson's approximation is

$$E(\Delta x) = \frac{b-a}{180}M(\Delta x)^4 = \frac{(b-a)^5}{180n^4}M.$$

Example 7.41: Approximate an Integral With Parabolas

Let us again approximate $\int_0^1 e^{-x^2}\,dx$ to two decimal places.

Solution. The fourth derivative of $f = e^{-x^2}$ is $(16x^4 - 48x^2 + 12)e^{-x^2}$; on $[0,1]$ this is at most 12 in absolute value. We begin by estimating the number of subintervals we are likely to need. To get two decimal places of accuracy, we will certainly need $E(\Delta x) < 0.005$, but taking a cue from our earlier example, let's require $E(\Delta x) < 0.001$:

$$\begin{aligned} \frac{1}{180}(12)\frac{1}{n^4} &< 0.001 \\ \frac{200}{3} &< n^4 \\ 2.86 \approx \sqrt[4]{\frac{200}{3}} &< n \end{aligned}$$

So we try $n = 4$, since we need an even number of subintervals. Then the error bound is $12/180/4^4 < 0.0003$ and the approximation is

$$(f(0)+4f(1/4)+2f(1/2)+4f(3/4)+f(1))\frac{1}{3\cdot 4} \approx 0.746855.$$

So the true value of the integral is between $0.746855 - 0.0003 = 0.746555$ and $0.746855 + 0.0003 = 0.7471555$, both of which round to 0.75. ♣

Exercises for 7.6

In the following problems, compute the trapezoid and Simpson approximations using 4 subintervals, and compute the error bound for each. (Finding the maximum values of the second and fourth derivatives can be challenging for some of these; you may use a graphing calculator or computer software to estimate the maximum values.)

Exercise 7.6.1 $\displaystyle\int_1^3 x\,dx$

Exercise 7.6.2 $\displaystyle\int_0^3 x^2\,dx$

Exercise 7.6.3 $\displaystyle\int_2^4 x^3\,dx$

Exercise 7.6.4 $\displaystyle\int_1^3 \frac{1}{x}\,dx$

Exercise 7.6.5 $\displaystyle\int_1^2 \frac{1}{1+x^2}\,dx$

Exercise 7.6.6 $\displaystyle\int_0^1 x\sqrt{1+x}\,dx$

Exercise 7.6.7 $\displaystyle\int_1^5 \frac{x}{1+x}\,dx$

Exercise 7.6.8 $\displaystyle\int_0^1 \sqrt{x^3+1}\,dx$

Exercise 7.6.9 $\displaystyle\int_0^1 \sqrt{x^4+1}\,dx$

Exercise 7.6.10 $\displaystyle\int_1^4 \sqrt{1+1/x}\,dx$

Exercise 7.6.11 *Using Simpson's rule on a parabola $f(x)$, even with just two subintervals, gives the exact value of the integral, because the parabolas used to approximate f will be f itself. Remarkably, Simpson's rule also computes the integral of a cubic function $f(x) = ax^3 + bx^2 + cx + d$ exactly. Show this is true by showing that*

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{x_2 - x_0}{3 \cdot 2}\left(f(x_0) + 4f((x_0 + x_2)/2) + f(x_2)\right).$$

This does require a bit of messy algebra, so you may prefer to use Sage.

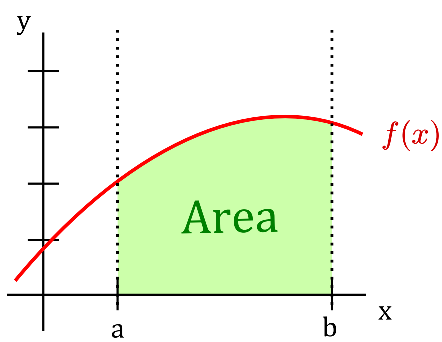
7.7 Improper Integrals

Recall that the Fundamental Theorem of Calculus says that if f is a **continuous** function on the closed interval $[a,b]$, then

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a),$$

where F is any antiderivative of f.

Both the **continuity** condition and **closed interval** must hold to use the Fundamental Theorem of Calculus, and in this case, $\int_a^b f(x)\,dx$ represents the net area under $f(x)$ from a to b:



We begin with an example where blindly applying the Fundamental Theorem of Calculus can give an incorrect result.

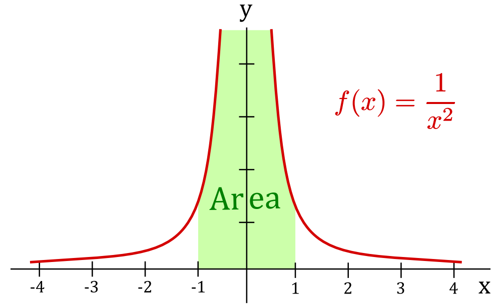
Example 7.42: Using FTC

Explain why $\int_{-1}^{1} \frac{1}{x^2}\,dx$ is not equal to -2.

Solution. Here is how one might proceed:

$$\int_{-1}^{1} \frac{1}{x^2}\,dx = \int_{-1}^{1} x^{-2}\,dx = -x^{-1}\Big|_{-1}^{1} = -\frac{1}{x}\Big|_{-1}^{1} = \left(-\frac{1}{1}\right) - \left(-\frac{1}{(-1)}\right) = -2$$

However, the above answer is **WRONG!** Since $f(x) = 1/x^2$ is not continuous on $[-1,\ 1]$, we cannot directly apply the Fundamental Theorem of Calculus. Intuitively, we can see why -2 is not the correct answer by looking at the graph of $f(x) = 1/x^2$ on $[-1,\ 1]$. The shaded area appears to grow without bound as seen in the figure below.



♣

Formalizing this example leads to the concept of an improper integral. There are two ways to extend the Fundamental Theorem of Calculus. One is to use an **infinite interval**, i.e., $[a,\infty)$, $(-\infty,b]$ or $(-\infty,\infty)$. The second is to allow the interval $[a,b]$ to contain an infinite **discontinuity** of $f(x)$. In either case, the integral is called an **improper integral**. One of the most important applications of this concept is probability distributions.

To compute improper integrals, we use the concept of limits along with the Fundamental Theorem of Calculus.

Definition 7.43: Definitions for Improper Integrals

If $f(x)$ is continuous on $[a,\infty)$, then the improper integral of f over $[a,\infty)$ is:

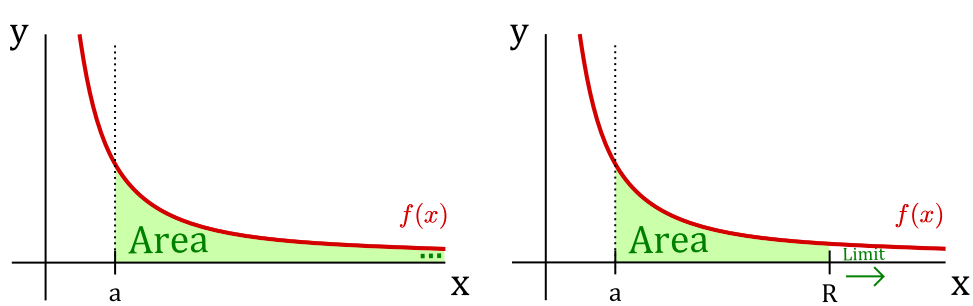
$$\int_a^\infty f(x)\,dx := \lim_{R\to\infty} \int_a^R f(x)\,dx.$$

If $f(x)$ is continuous on $(-\infty,b]$, then the improper integral of f over $(-\infty,b]$ is:

$$\int_{-\infty}^b f(x)\,dx := \lim_{R\to-\infty} \int_R^b f(x)\,dx.$$

If the limit exists and is a finite number, we say the improper integral **converges**. Otherwise, we say the improper integral **diverges**.

To get an intuitive (though not completely correct) interpretation of improper integrals, we attempt to analyze $\int_a^\infty f(x)\,dx$ graphically. Here assume $f(x)$ is continuous on $[a,\infty)$:



We let R be a fixed number in $[a,\infty)$. Then by taking the limit as R approaches ∞, we get the improper integral:

$$\int_a^\infty f(x)\,dx := \lim_{R\to\infty}\int_a^R f(x)\,dx.$$

We can then apply the Fundamental Theorem of Calculus to the last integral as $f(x)$ is continuous on the closed interval $[a,R]$.

We next define the improper integral for the interval $(-\infty,\ \infty)$.

Definition 7.44: Definitions for Improper Integrals

If both $\int_{-\infty}^a f(x)\,dx$ *and* $\int_a^\infty f(x)\,dx$ *are convergent, then the improper integral of* f *over* $(-\infty,\infty)$ *is:*

$$\int_{-\infty}^\infty f(x)\,dx := \int_{-\infty}^a f(x)\,dx + \int_a^\infty f(x)\,dx$$

The above definition requires **both** of the integrals

$$\int_{-\infty}^a f(x)\,dx \qquad \text{and} \qquad \int_a^\infty f(x)\,dx$$

to be convergent for $\int_{-\infty}^\infty f(x)\,dx$ to also be convergent. If **either** of $\int_{-\infty}^a f(x)\,dx$ or $\int_a^\infty f(x)\,dx$ is divergent, then so is $\int_{-\infty}^\infty f(x)\,dx$.

Example 7.45: Improper Integral

Determine whether $\int_1^\infty \frac{1}{x}\,dx$ *is convergent or divergent.*

Solution. Using the definition for improper integrals we write this as:

$$\int_1^\infty \frac{1}{x}\,dx = \lim_{R\to\infty}\int_1^R \frac{1}{x}\,dx = \lim_{R\to\infty}\ln|x|\Big|_1^R = \lim_{R\to\infty}\ln|R| - \ln|1| = \lim_{R\to\infty}\ln|R| = +\infty$$

Therefore, the integral is **divergent**. ♣

Example 7.46: Improper Integral

Determine whether $\int_{-\infty}^\infty x\sin(x^2)\,dx$ *is convergent or divergent.*

Solution. We must compute both $\int_0^\infty x\sin(x^2)\,dx$ and $\int_{-\infty}^0 x\sin(x^2)\,dx$. Note that we don't have to split the integral up at 0, any finite value a will work. First we compute the indefinite integral. Let $u = x^2$, then

$du = 2x\,dx$ and hence,

$$\int x\sin(x^2)\,dx = \frac{1}{2}\int \sin(u)\,du = -\frac{1}{2}\cos(x^2) + C$$

Using the definition of improper integral gives:

$$\int_0^\infty x\sin(x^2)\,dx = \lim_{R\to\infty}\int_0^R x\sin(x^2)\,dx = \lim_{R\to\infty}\left[-\frac{1}{2}\cos(x^2)\right]\Bigg|_0^R = -\frac{1}{2}\lim_{R\to\infty}\cos(R^2) + \frac{1}{2}$$

This limit does not exist since $\cos x$ **oscillates** between -1 and $+1$. In particular, $\cos x$ does not approach any particular value as x gets larger and larger. Thus, $\int_0^\infty x\sin(x^2)\,dx$ diverges, and hence, $\int_{-\infty}^\infty x\sin(x^2)\,dx$ diverges. ♣

When there is a discontinuity in $[a,b]$ or at an endpoint, then the improper integral is as follows.

Definition 7.47: Definitions for Improper Integrals

If $f(x)$ is continuous on $(a,b]$, then the improper integral of f over $(a,b]$ is:

$$\int_a^b f(x)\,dx := \lim_{R\to a^+}\int_R^b f(x)\,dx.$$

If $f(x)$ is continuous on $[a,b)$, then the improper integral of f over $[a,b)$ is:

$$\int_a^b f(x)\,dx := \lim_{R\to b^-}\int_a^R f(x)\,dx.$$

If the limit above exists and is a finite number, we say the improper integral **converges**. Otherwise, we say the improper integral **diverges**.

When there is a discontinuity in the interior of $[a,b]$, we use the following definition.

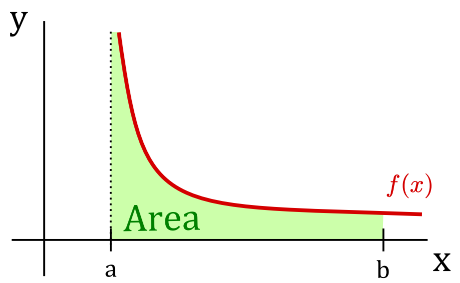
Definition 7.48: Definitions for Improper Integrals

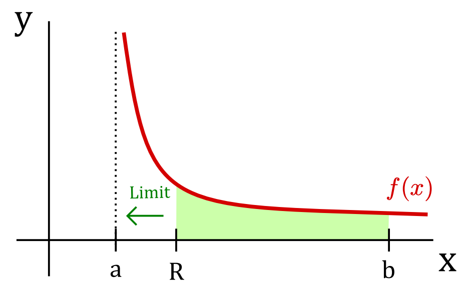
If f has a discontinuity at $x = c$ where $c \in [a,b]$, and both both $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ are convergent, then f over $[a,b]$ is:

$$\int_a^b f(x)\,dx := \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

Again, we can get an intuitive sense of this concept by analyzing $\int_a^b f(x)\,dx$ graphically. Here assume

$f(x)$ is continuous on $(a,b]$ but discontinuous at $x=a$:





We let R be a fixed number in (a,b). Then by taking the limit as R approaches a from the **right**, we get the improper integral:

$$\int_a^b f(x)\,dx := \lim_{R\to a^+}\int_R^b f(x)\,dx.$$

Now we can apply FTC to the last integral as $f(x)$ is continuous on $[R,b]$.

Example 7.49: A Divergent Integral

Determine if $\int_{-1}^{1}\frac{1}{x^2}\,dx$ *is convergent or divergent.*

Solution. The function $f(x)=1/x^2$ has a discontinuity at $x=0$, which lies in $[-1,1]$. We must compute $\int_{-1}^{0}\frac{1}{x^2}\,dx$ and $\int_{0}^{1}\frac{1}{x^2}\,dx$. Let's start with $\int_{0}^{1}\frac{1}{x^2}\,dx$:

$$\int_0^1\frac{1}{x^2}\,dx = \lim_{R\to 0^+}\int_R^1\frac{1}{x^2}\,dx = \lim_{R\to 0^+} -\frac{1}{x}\bigg|_R^1 = -1+\lim_{R\to 0^+}\frac{1}{R}$$

which diverges to $+\infty$. Therefore, $\int_{-1}^{1}\frac{1}{x^2}\,dx$ is **divergent** since one of $\int_{-1}^{0}\frac{1}{x^2}\,dx$ and $\int_{0}^{1}\frac{1}{x^2}\,dx$ is divergent.

♣

Example 7.50: Integral of the Logarithm

Determine if $\int_0^1 \ln x\,dx$ *is convergent or divergent. Evaluate it if it is convergent.*

Solution. Note that $f(x)=\ln x$ is discontinuous at the endpoint $x=0$. We first use integration by parts to compute $\int \ln x\,dx$. We let $u=\ln x$ and $dv=dx$. Then $du=(1/x)dx$, $v=x$, giving:

$$\begin{aligned}\int \ln x\,dx &= x\ln x-\int x\cdot\frac{1}{x}\,dx\\ &= x\ln x-\int 1\,dx\\ &= x\ln x-x+C\end{aligned}$$

Now using the definition of improper integral for $\int_0^1 \ln x \, dx$:

$$\int_0^1 \ln x \, dx = \lim_{R \to 0^+} \int_R^1 \ln x \, dx = \lim_{R \to 0^+} (x \ln x - x)\Big|_R^1 = -1 - \lim_{R \to 0^+} (R \ln R) + \lim_{R \to 0^+} R$$

Note that $\lim_{R \to 0^+} R = 0$. We next compute $\lim_{R \to 0^+} (R \ln R)$. First, we rewrite the expression as follows:

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

Now the limit is of the indeterminate type $(-\infty)/(\infty)$ and l'Hôpital's Rule can be applied.

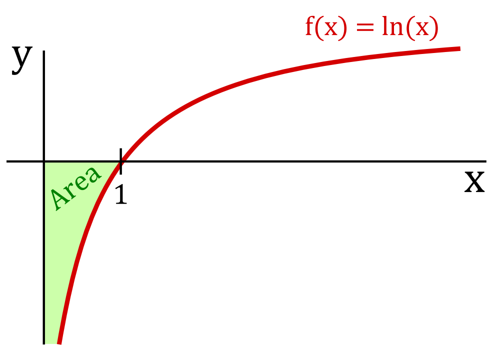
$$\lim_{R \to 0^+} (R \ln R) = \lim_{R \to 0^+} \frac{\ln R}{1/R} = \lim_{R \to 0^+} \frac{1/R}{-1/R^2} = \lim_{R \to 0^+} -\frac{R^2}{R} = \lim_{R \to 0^+} (-R) = 0$$

Thus, $\lim_{R \to 0^+} (R \ln R) = 0$. Thus

$$\int_0^1 \ln x \, dx = -1,$$

and the integral is convergent to -1.

Graphically, one might interpret this to mean that the net area under $\ln x$ on $[0,1]$ is -1 (the area in this case lies below the x-axis).



♣

Example 7.51: Integral of a Square Root

Determine if $\int_0^4 \frac{dx}{\sqrt{4-x}}$ *is convergent or divergent. Evaluate it if it is convergent.*

Solution. Note that $\frac{1}{\sqrt{4-x}}$ is discontinuous at the endpoint $x = 4$. We use a u-substitution to compute $\int \frac{dx}{\sqrt{4-x}}$. We let $u = 4 - x$, then $du = -dx$, giving:

$$\int \frac{dx}{\sqrt{4-x}} = \int -\frac{du}{u^{1/2}}$$

$$\begin{aligned} &= \int -u^{-1/2}\,du \\ &= -2(u)^{1/2} + C \\ &= -2\sqrt{4-x} + C \end{aligned}$$

Now using the definition of improper integrals for $\displaystyle\int_0^4 \frac{dx}{\sqrt{4-x}}$:

$$\int_0^4 \frac{dx}{\sqrt{4-x}} = \lim_{R\to 4^-} (-2\sqrt{4-x})\Big|_0^R = \lim_{R\to 4^-} -2\sqrt{4-R} + 2\sqrt{4} = 4$$

♣

Example 7.52: Improper Integral

Determine if $\displaystyle\int_1^2 \frac{dx}{(x-1)^{1/3}}$ *is convergent or divergent. Evaluate it if it is convergent.*

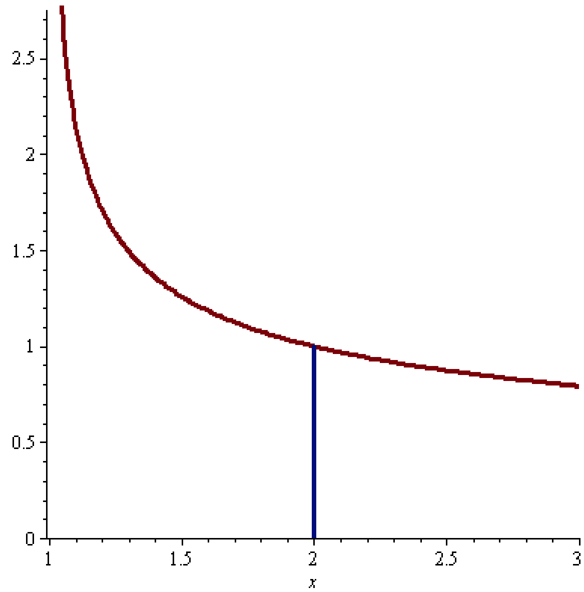
Solution. Note that $f(x) = \dfrac{1}{(x-1)^{1/3}}$ is discontinuous at the endpoint $x = 1$. We first use substitution to find $\displaystyle\int \frac{dx}{(x-1)^{1/3}}$. We let $u = x-1$. Then $du = dx$, giving

$$\int \frac{dx}{(x-1)^{1/3}} = \int \frac{du}{u^{1/3}} = \int u^{-1/3}du = \frac{3}{2}u^{2/3} + C = \frac{3}{2}(x-1)^{2/3} + C.$$

Now using the definition of improper integral for $\displaystyle\int_1^2 \frac{dx}{(x-1)^{1/3}}$:

$$\int_1^2 \frac{dx}{(x-1)^{1/3}} = \lim_{R\to 1^+} \int_R^2 \frac{dx}{(x-1)^{1/3}} = \lim_{R\to 1^+} \frac{3}{2}(x-1)^{2/3}\Big|_R^2 = \frac{3}{2} - \lim_{R\to 1^+} \frac{3}{2}(R-1)^{2/3} = \frac{3}{2},$$

and the integral is convergent to $\frac{3}{2}$. Graphically, one might interpret this to mean that the net area under $\dfrac{1}{(x-1)^{1/3}}$ on $[1,2]$ is $\frac{3}{2}$.



♣

The following test allows us to determine convergence/divergence information about improper integrals that are hard to compute by comparing them to easier ones. We state the test for $[a,\infty)$, but similar versions hold for the other improper integrals.

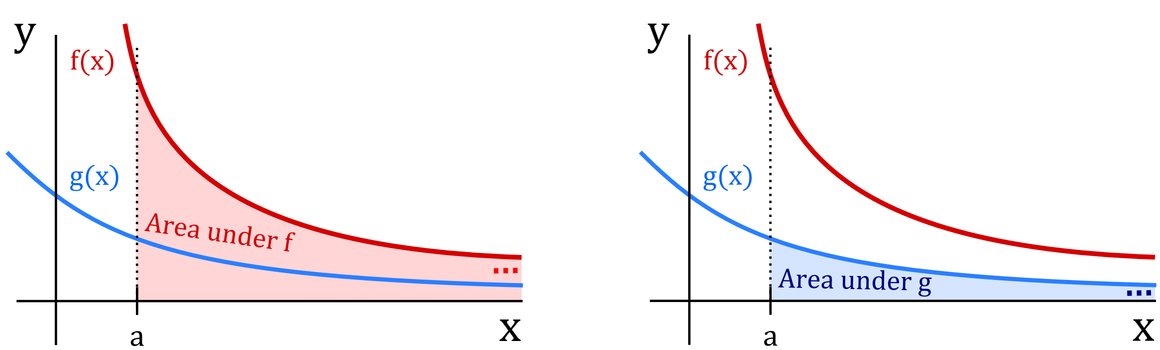
The Comparison Test

Assume that $f(x) \geq g(x) \geq 0$ for $x \geq a$.

(i) If $\displaystyle\int_a^{\infty} f(x)\,dx$ **converges**, then $\displaystyle\int_a^{\infty} g(x)\,dx$ also **converges**.

(ii) If $\displaystyle\int_a^{\infty} g(x)\,dx$ **diverges**, then $\displaystyle\int_a^{\infty} f(x)\,dx$ also **diverges**.

Informally, (i) says that if $f(x)$ is larger than $g(x)$, and the area under $f(x)$ is finite (converges), then the area under $g(x)$ must also be finite (converges). Informally, (ii) says that if $f(x)$ is larger than $g(x)$, and the area under $g(x)$ is infinite (diverges), then the area under $f(x)$ must also be infinite (diverges).



Example 7.53: Comparison Test

Show that $\int_2^\infty \frac{\cos^2 x}{x^2}\,dx$ *converges.*

Solution. We use the comparison test to show that it converges. Note that $0 \le \cos^2 x \le 1$ and hence

$$0 \le \frac{\cos^2 x}{x^2} \le \frac{1}{x^2}.$$

Thus, taking $f(x) = 1/x^2$ and $g(x) = \cos^2 x/x^2$ we have $f(x) \ge g(x) \ge 0$. One can easily see that $\int_2^\infty \frac{1}{x^2}\,dx$ converges. Therefore, $\int_2^\infty \frac{\cos^2 x}{x^2}\,dx$ also converges. ♣

Exercises for Section 7.7

Exercise 7.7.1 *Determine whether* $\int_1^\infty \frac{1}{x^2}\,dx$ *is convergent or divergent.*

Exercise 7.7.2 *Determine whether* $\int_e^\infty \frac{1}{x\sqrt{\ln x}}\,dx$ *is convergent or divergent.*

Exercise 7.7.3 *Evaluate the improper integral* $\int_0^\infty e^{-3x}\,dx$.

Exercise 7.7.4 *Determine if* $\int_1^e \frac{1}{x(\ln x)^2}\,dx$ *is convergent or divergent. Evaluate it if it is convergent.*

Exercise 7.7.5 *Show that* $\int_0^\infty e^{-x}\sin^2\left(\frac{\pi x}{2}\right)dx$ *converges.*

Exercise 7.7.6 *Evaluate* $\int_{-\infty}^\infty \frac{1}{x^2+1}\,dx$ *and* $\int_{-\infty}^\infty \frac{x}{x^2+1}\,dx$.

Exercise 7.7.7 *Determine whether the following improper integrals are convergent or divergent. Evaluate those that are convergent.*

(a) $\int_0^\infty \frac{1}{x^2+1}\,dx$

(b) $\int_0^\infty \frac{x}{x^2+1}\,dx$

(c) $\int_0^\infty e^{-x}(\cos x + \sin x)\,dx$. *[Hint: What is the derivative of* $-e^{-x}\cos x$*?]*

(d) $\int_0^{\pi/2} \sec^2 x\,dx$

(e) $\int_0^4 \dfrac{1}{(4-x)^{2/5}}\,dx$

Exercise 7.7.8 *Prove that the integral* $\int_1^\infty \dfrac{1}{x^p}\,dx$ *is convergent if* $p > 1$ *and divergent if* $0 < p \leq 1$.

Exercise 7.7.9 *Suppose that* $p > 0$. *Find all values of* p *for which* $\int_0^1 \dfrac{1}{x^p}\,dx$ *converges.*

Exercise 7.7.10 *Show that* $\int_1^\infty \dfrac{\sin^2 x}{x(\sqrt{x}+1)}\,dx$ *converges.*

7.8 Additional exercises

These problems require the techniques of this chapter, and are in no particular order. Some problems may be done in more than one way.

Exercise 7.8.1 $\displaystyle\int (t+4)^3\,dt$

Exercise 7.8.2 $\displaystyle\int t(t^2-9)^{3/2}\,dt$

Exercise 7.8.3 $\displaystyle\int (e^{t^2}+16)te^{t^2}\,dt$

Exercise 7.8.4 $\displaystyle\int \sin t\cos 2t\,dt$

Exercise 7.8.5 $\displaystyle\int \tan t\sec^2 t\,dt$

Exercise 7.8.6 $\displaystyle\int \frac{2t+1}{t^2+t+3}\,dt$

Exercise 7.8.7 $\displaystyle\int \frac{1}{t(t^2-4)}\,dt$

Exercise 7.8.8 $\displaystyle\int \frac{1}{(25-t^2)^{3/2}}\,dt$

Exercise 7.8.9 $\displaystyle\int \frac{\cos 3t}{\sqrt{\sin 3t}}\,dt$

Exercise 7.8.10 $\int t\sec^2 t\,dt$

Exercise 7.8.11 $\int \frac{e^t}{\sqrt{e^t+1}}\,dt$

Exercise 7.8.12 $\int \cos^4 t\,dt$

Exercise 7.8.13 $\int \frac{1}{t^2+3t}\,dt$

Exercise 7.8.14 $\int \frac{1}{t^2\sqrt{1+t^2}}\,dt$

Exercise 7.8.15 $\int \frac{\sec^2 t}{(1+\tan t)^3}\,dt$

Exercise 7.8.16 $\int t^3\sqrt{t^2+1}\,dt$

Exercise 7.8.17 $\int e^t\sin t\,dt$

Exercise 7.8.18 $\int (t^{3/2}+47)^3\sqrt{t}\,dt$

Exercise 7.8.19 $\int \frac{t^3}{(2-t^2)^{5/2}}\,dt$

Exercise 7.8.20 $\int \frac{1}{t(9+4t^2)}\,dt$

Exercise 7.8.21 $\int \frac{\arctan 2t}{1+4t^2}\,dt$

Exercise 7.8.22 $\int \frac{t}{t^2+2t-3}\,dt$

Exercise 7.8.23 $\int \sin^3 t\cos^4 t\,dt$

Exercise 7.8.24 $\int \frac{1}{t^2-6t+9}\,dt$

Exercise 7.8.25 $\int \frac{1}{t(\ln t)^2}\,dt$

Exercise 7.8.26 $\int t(\ln t)^2\,dt$

Exercise 7.8.27 $\int t^3 e^t \, dt$

Exercise 7.8.28 $\int \frac{t+1}{t^2+t-1} \, dt$

8. Applications of Integration

8.1 Distance, Velocity, Acceleration

We recall a general principle that will later be applied to distance-velocity-acceleration problems, among other things. If $F(u)$ is an anti-derivative of $f(u)$, then $\int_a^b f(u)\,du = F(b) - F(a)$. Suppose that we want to let the upper limit of integration vary, i.e., we replace b by some variable x. We think of a as a fixed starting value x_0. In this new notation the last equation (after adding $F(a)$ to both sides) becomes:

$$F(x) = F(x_0) + \int_{x_0}^{x} f(u)\,du.$$

Here u is the variable of integration, called a "dummy variable," since it is not the variable in the function $F(x)$. In general, it is not a good idea to use the same letter as a variable of integration and as a limit of integration. That is, $\int_{x_0}^{x} f(x)dx$ is bad notation, and can lead to errors and confusion.

An important application of this principle occurs when we are interested in the position of an object at time t (say, on the x-axis) and we know its position at time t_0. Let $s(t)$ denote the position of the object at time t (its distance from a reference point, such as the origin on the x-axis). Then the net change in position between t_0 and t is $s(t) - s(t_0)$. Since $s(t)$ is an anti-derivative of the velocity function $v(t)$, we can write

$$s(t) = s(t_0) + \int_{t_0}^{t} v(u)du.$$

Similarly, since the velocity is an anti-derivative of the acceleration function $a(t)$, we have

$$v(t) = v(t_0) + \int_{t_0}^{t} a(u)du.$$

Example 8.1: Constant Force

Suppose an object is acted upon by a constant force F. Find $v(t)$ and $s(t)$.

Solution. By Newton's law $F = ma$, so the acceleration is F/m, where m is the mass of the object. Then we first have

$$v(t) = v(t_0) + \int_{t_0}^{t} \frac{F}{m}\,du = v_0 + \left.\frac{F}{m}u\right|_{t_0}^{t} = v_0 + \frac{F}{m}(t - t_0),$$

using the usual convention $v_0 = v(t_0)$. Then

$$s(t) = s(t_0) + \int_{t_0}^{t} \left(v_0 + \tfrac{F}{m}(u - t_0)\right) du = s_0 + \left.\left(v_0 u + \tfrac{F}{2m}(u - t_0)^2\right)\right|_{t_0}^{t}$$

$$= s_0 + v_0(t - t_0) + \tfrac{F}{2m}(t - t_0)^2.$$

For instance, when $F/m = -g$ is the constant of gravitational acceleration, then this is the falling body formula (if we neglect air resistance) familiar from elementary physics:

$$s_0 + v_0(t - t_0) - \frac{g}{2}(t - t_0)^2,$$

or in the common case that $t_0 = 0$,

$$s_0 + v_0 t - \frac{g}{2}t^2.$$

♣

Recall that the integral of the velocity function gives the *net* distance traveled. If you want to know the *total* distance traveled, you must find out where the velocity function crosses the t-axis, integrate separately over the time intervals when $v(t)$ is positive and when $v(t)$ is negative, and add up the absolute values of the different integrals. For example, if an object is thrown straight upward at 19.6 m/sec, its velocity function is $v(t) = -9.8t + 19.6$, using $g = 9.8$ m/sec for the force of gravity. This is a straight line which is positive for $t < 2$ and negative for $t > 2$. The net distance traveled in the first 4 seconds is thus

$$\int_0^4 (-9.8t + 19.6)dt = 0,$$

while the total distance traveled in the first 4 seconds is

$$\int_0^2 (-9.8t + 19.6)dt + \left|\int_2^4 (-9.8t + 19.6)dt\right| = 19.6 + |-19.6| = 39.2$$

meters, 19.6 meters up and 19.6 meters down.

Example 8.2: Net and Total Distance

The acceleration of an object is given by $a(t) = \cos(\pi t)$, and its velocity at time $t = 0$ is $1/(2\pi)$. Find both the net and the total distance traveled in the first 1.5 seconds.

Solution. We compute

$$v(t) = v(0) + \int_0^t \cos(\pi u)du = \frac{1}{2\pi} + \frac{1}{\pi}\sin(\pi u)\Big|_0^t = \frac{1}{\pi}\left(\frac{1}{2} + \sin(\pi t)\right).$$

The *net* distance traveled is then

$$\begin{aligned} s(3/2) - s(0) &= \int_0^{3/2} \frac{1}{\pi}\left(\frac{1}{2} + \sin(\pi t)\right) dt \\ &= \frac{1}{\pi}\left(\frac{t}{2} - \frac{1}{\pi}\cos(\pi t)\right)\Big|_0^{3/2} \\ &= \frac{3}{4\pi} + \frac{1}{\pi^2} \approx 0.340 \text{ meters.} \end{aligned}$$

To find the *total* distance traveled, we need to know when $(0.5+\sin(\pi t))$ is positive and when it is negative. This function is 0 when $\sin(\pi t)$ is -0.5, i.e., when $\pi t = 7\pi/6$, $11\pi/6$, etc. The value $\pi t = 7\pi/6$, i.e., $t = 7/6$, is the only value in the range $0 \le t \le 1.5$. Since $v(t) > 0$ for $t < 7/6$ and $v(t) < 0$ for $t > 7/6$, the total distance traveled is

$$\begin{aligned}
&\int_0^{7/6} \frac{1}{\pi}\left(\frac{1}{2}+\sin(\pi t)\right)dt + \left|\int_{7/6}^{3/2} \frac{1}{\pi}\left(\frac{1}{2}+\sin(\pi t)\right)dt\right| \\
&= \frac{1}{\pi}\left(\frac{7}{12}+\frac{1}{\pi}\cos(7\pi/6)+\frac{1}{\pi}\right)+\frac{1}{\pi}\left|\frac{3}{4}-\frac{7}{12}+\frac{1}{\pi}\cos(7\pi/6)\right| \\
&= \frac{1}{\pi}\left(\frac{7}{12}+\frac{1}{\pi}\frac{\sqrt{3}}{2}+\frac{1}{\pi}\right)+\frac{1}{\pi}\left|\frac{3}{4}-\frac{7}{12}+\frac{1}{\pi}\frac{\sqrt{3}}{2}.\right| \\
&\approx 0.409 \text{ meters.}
\end{aligned}$$

♣

Exercises for Section 8.1

Exercise 8.1.1 *An object moves so that its velocity at time t is $v(t) = -9.8t + 20$ m/s. Describe the motion of the object between $t = 0$ and $t = 5$, find the total distance traveled by the object during that time, and find the net distance traveled.*

Exercise 8.1.2 *An object moves so that its velocity at time t is $v(t) = \sin t$. Set up and evaluate a single definite integral to compute the net distance traveled between $t = 0$ and $t = 2\pi$.*

Exercise 8.1.3 *An object moves so that its velocity at time t is $v(t) = 1 + 2\sin t$ m/s. Find the net distance traveled by the object between $t = 0$ and $t = 2\pi$, and find the total distance traveled during the same period.*

Exercise 8.1.4 *Consider the function $f(x) = (x+2)(x+1)(x-1)(x-2)$ on $[-2,2]$. Find the total area between the curve and the x-axis (measuring all area as positive).*

Exercise 8.1.5 *Consider the function $f(x) = x^2 - 3x + 2$ on $[0,4]$. Find the total area between the curve and the x-axis (measuring all area as positive).*

Exercise 8.1.6 *Evaluate the three integrals:*

$$A = \int_0^3 (-x^2+9)\,dx \qquad B = \int_0^4 (-x^2+9)\,dx \qquad C = \int_4^3 (-x^2+9)\,dx,$$

and verify that $A = B + C$.

8.2 Area Between Curves

We have seen how integration can be used to find an area between a curve and the x-axis. With very little change we can find some areas between curves; indeed, the area between a curve and the x-axis may be interpreted as the area between the curve and a second "curve" with equation $y = 0$.

Suppose we would like to find the area below $f(x) = -x^2 + 4x + 3$ and above $g(x) = -x^3 + 7x^2 - 10x + 5$ over the interval $1 \le x \le 2$. We can approximate the area between two curves by dividing the area into thin sections and approximating the area of each section by a rectangle, as indicated in figure 8.1. The area of a typical rectangle is $\Delta x(f(x_i) - g(x_i))$, so the total area is approximately

$$\sum_{i=0}^{n-1} (f(x_i) - g(x_i))\Delta x.$$

This is exactly the sort of sum that turns into an integral in the limit, namely the integral

$$\int_1^2 f(x) - g(x)\, dx.$$

Then

$$\int_1^2 f(x) - g(x)\, dx = \int_1^2 (-x^2 + 4x + 3) - (-x^3 + 7x^2 - 10x + 5)\, dx = \frac{49}{12}.$$

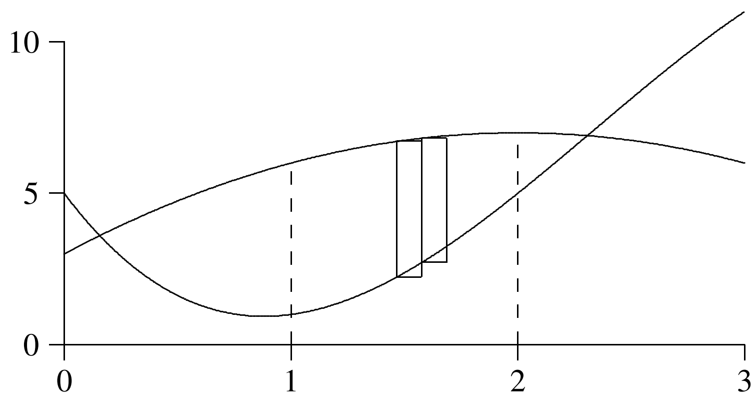
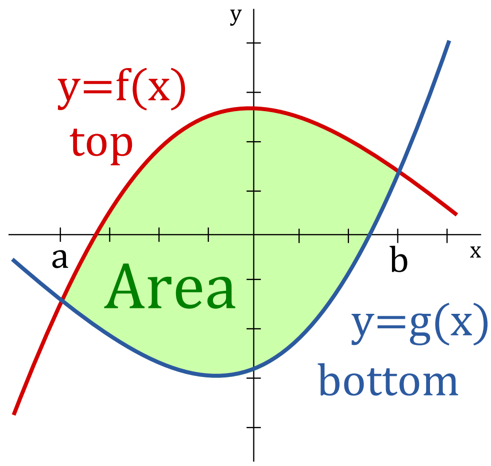


Figure 8.1: Approximating area between curves with rectangles.

This procedure can informally be thought of as follows.

Area Between Two Curves

$$Area = \int_a^b (\text{top curve}) - (\text{bottom curve})\, dx, \qquad a \le x \le b.$$



More formally, the area A of the region bounded by the curves $y = f(x)$ and $y = g(x)$ and the lines $x = a$ and $x = b$ is:

$$A = \int_a^b |f(x) - g(x)|\, dx.$$

Example 8.3: Area between Curves

Find the area between $f(x) = -x^2 + 4x$ and $g(x) = x^2 - 6x + 5$; the curves are shown in figure 8.2.

Solution. Here we are not given a specific interval, so it must be the case that there is a "natural" region involved. Since the curves are both parabolas, the only reasonable interpretation is the region between the two intersection points, which can be computed as:

$$\frac{5 \pm \sqrt{15}}{2}.$$

If we let $a = (5 - \sqrt{15})/2$ and $b = (5 + \sqrt{15})/2$, the total area is

$$\begin{aligned} \int_a^b -x^2 + 4x - (x^2 - 6x + 5)\, dx &= \int_a^b -2x^2 + 10x - 5\, dx \\ &= -\frac{2x^3}{3} + 5x^2 - 5x\Big|_a^b \\ &= 5\sqrt{15}. \end{aligned}$$

after a bit of simplification. ♣

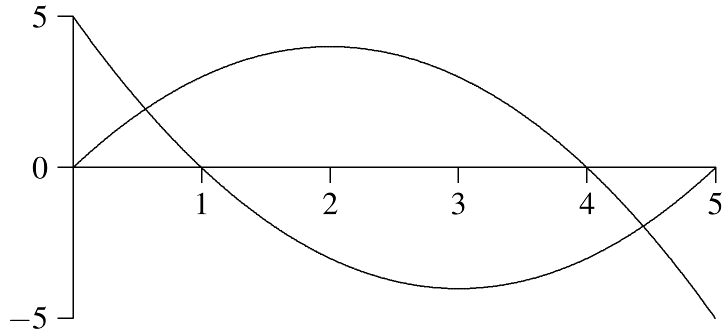


Figure 8.2: Area bounded by two curves.

Some general guidelines to compute the area between two curves follows.

Guidelines for Area Between Two Curves

1. Find the intersection points.
2. Draw a sketch of the two curves.
3. Using the sketch determine which curve is the top curve and which curve is the bottom curve. You may need to split the area up into multiple regions if the curves intersect multiple times in $[a,b]$.
4. Put the above information into the appropriate formula (once for each region):

$$\textbf{Area} = \int_a^b (\text{top curve}) - (\text{bottom curve})\, dx, \qquad a \le x \le b.$$

5. Evaluate the integral using the Fundamental Theorem of Calculus (you should get a positive number representing an area).

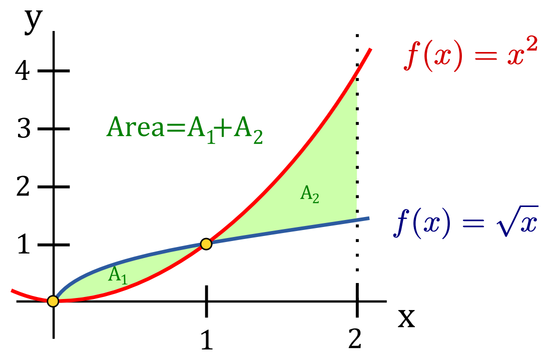
Example 8.4: Area Between Two Curves

Determine the area enclosed by $y = x^2$, $y = \sqrt{x}$, $x = 0$ *and* $x = 2$.

Solution. The points of intersection of $y = x^2$ and $y = \sqrt{x}$ are

$$x^2 = \sqrt{x} \qquad \to \qquad x^4 = x \qquad \to \qquad x^4 - x = 0 \qquad \to \qquad x(x^3 - 1) = 0.$$

Thus, either $x = 0$ or $x = 1$. Sketching the curves gives:



The area we want to compute is the shaded region. Since the top curve changes at $x = 1$, we need to use the formula twice. For A_1 we have $a = 0$, $b = 1$, the top curve is $y = \sqrt{x}$ and the bottom curve is $y = x^2$. For A_2 we have $a = 1$, $b = 2$, the top curve is $y = x^2$ and the bottom curve is $y = \sqrt{x}$.

$$\textbf{Area} = \mathbf{A}_1 + \mathbf{A}_2 = \int_0^1 (\sqrt{x} - x^2)\,dx + \int_1^2 (x^2 - \sqrt{x})\,dx$$

For the first integral we have:

$$\int_0^1 (\sqrt{x} - x^2)\,dx = \left(\frac{2}{3}x^{3/2} - \frac{1}{3}x^3\right)\Bigg|_0^1 = \frac{1}{3}$$

Thus,

$$\textbf{Area} = \frac{1}{3} + \left(\frac{1}{3}x^3 - \frac{2}{3}x^{3/2}\right)\Bigg|_1^2 = \frac{1}{3} + \left[\left(\frac{8}{3} - \frac{2(\sqrt{2})^3}{3}\right) - \left(\frac{1}{3} - \frac{2}{3}\right)\right] = \frac{10 - 4\sqrt{2}}{3}$$

♣

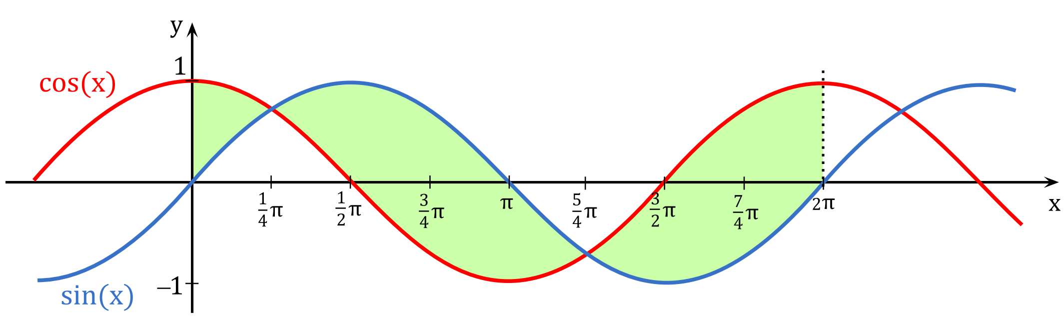
Example 8.5: Area Between Sine and Cosine

Determine the area enclosed by $y = \sin x$ *and* $y = \cos x$ *on the interval* $[0,\ 2\pi]$.

Solution. The curves $y = \sin x$ and $y = \cos x$ intersect when:

$$\sin x = \cos x \quad \rightarrow \quad \tan x = 1 \quad \rightarrow \quad x = \frac{\pi}{4} + \pi k,\ k \text{ an integer.}$$

We have the following sketch:

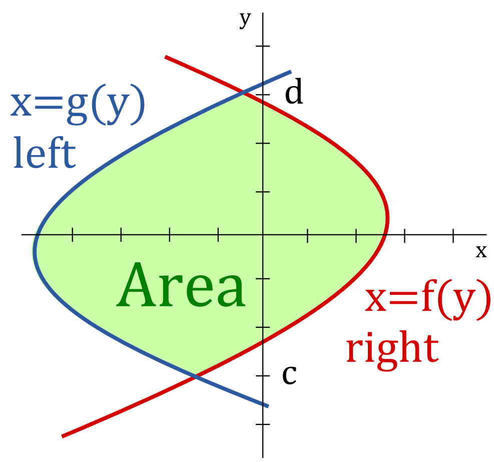


The area we want to compute is the shaded region. The top curve changes at $x = \pi/4$ and $x = 5\pi/4$, thus, we need to split the area up into three regions: from 0 to $\pi/4$; from $\pi/4$ to $5\pi/4$; and from $5\pi/4$ to 2π.

$$\begin{aligned}
\textbf{Area} &= \int_0^{\frac{\pi}{4}} (\cos x - \sin x)\,dx + \int_{\frac{\pi}{4}}^{\frac{5\pi}{4}} (\sin x - \cos x)\,dx + \int_{\frac{5\pi}{4}}^{2\pi} (\cos x - \sin x)\,dx \\
&= (\sin x + \cos x)\Big|_0^{\pi/4} + (-\cos x - \sin x)\Big|_{\pi/4}^{5\pi/4} + (\sin x + \cos x)\Big|_{5\pi/4}^{2\pi} \\
&= \left(\sqrt{2}-1\right) + \left(\sqrt{2}+\sqrt{2}\right) + \left(1+\sqrt{2}\right) \\
&= 4\sqrt{2}
\end{aligned}$$

♣

Sometimes the given curves are not functions of x. In this instances, it may be more useful to use the following.



The area A of the region bounded by the curves $x = f(y)$ and $x = g(y)$ and the lines $y = c$ and $y = d$ is:

$$A = \int_c^d |f(y) - g(y)|\,dy.$$

Informally this can be thought of as follows:

Area Between Two Curves

$$Area = \int_c^d (\text{right curve}) - (\text{left curve})\,dy, \qquad c \le y \le d.$$

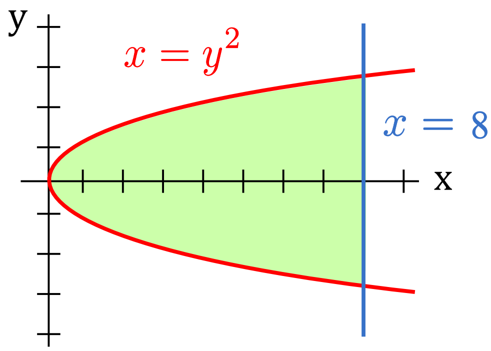
Example 8.6: Area Between Two Curves

Determine the area enclosed by $x = y^2$ and $x = 8$.

Solution. Note that $x = y^2$ and $x = 8$ intersect when:

$$y^2 = 8 \qquad \rightarrow \qquad y = \pm\sqrt{8} \qquad \rightarrow \qquad y = \pm 2\sqrt{2}$$

Sketching the two curves gives:



From the sketch $c = -2\sqrt{2}$, $d = 2\sqrt{2}$, the right curve is $x = 8$ and the left curve is $x = y^2$.

$$\textbf{Area} = \int_c^d [\text{right} - \text{left}]\, dy = \int_{-2\sqrt{2}}^{2\sqrt{2}} (8 - y^2)\, dy = \left(8y - \frac{1}{3}y^3\right)\Big|_{-2\sqrt{2}}^{2\sqrt{2}}$$

$$= \left[8(2\sqrt{2}) - \frac{1}{3}(2\sqrt{2})^3\right] - \left[8(-2\sqrt{2}) - \frac{1}{3}(-2\sqrt{2})^3\right] = \frac{64\sqrt{2}}{3}$$

♣

Exercises for Section 8.2

Find the area bounded by the curves.

Exercise 8.2.1 $y = x^4 - x^2$ *and* $y = x^2$ *(the part to the right of the y-axis)*

Exercise 8.2.2 $x = y^3$ *and* $x = y^2$

Exercise 8.2.3 $x = 1 - y^2$ *and* $y = -x - 1$

Exercise 8.2.4 $x = 3y - y^2$ *and* $x + y = 3$

Exercise 8.2.5 $y = \cos(\pi x/2)$ *and* $y = 1 - x^2$ *(in the first quadrant)*

Exercise 8.2.6 $y = \sin(\pi x/3)$ *and* $y = x$ *(in the first quadrant)*

Exercise 8.2.7 $y = \sqrt{x}$ *and* $y = x^2$

Exercise 8.2.8 $y = \sqrt{x}$ *and* $y = \sqrt{x+1}$, $0 \le x \le 4$

Exercise 8.2.9 $x = 0$ *and* $x = 25 - y^2$

Exercise 8.2.10 $y = \sin x \cos x$ *and* $y = \sin x$, $0 \le x \le \pi$

Exercise 8.2.11 $y = x^{3/2}$ *and* $y = x^{2/3}$

Exercise 8.2.12 $y = x^2 - 2x$ *and* $y = x - 2$

8.3 Volume

Now that we have seen how to compute certain areas by using integration; we will now look into how some volumes may also be computed by evaluating an integral. Generally, the volumes that we can compute this way have cross-sections that are easy to describe.

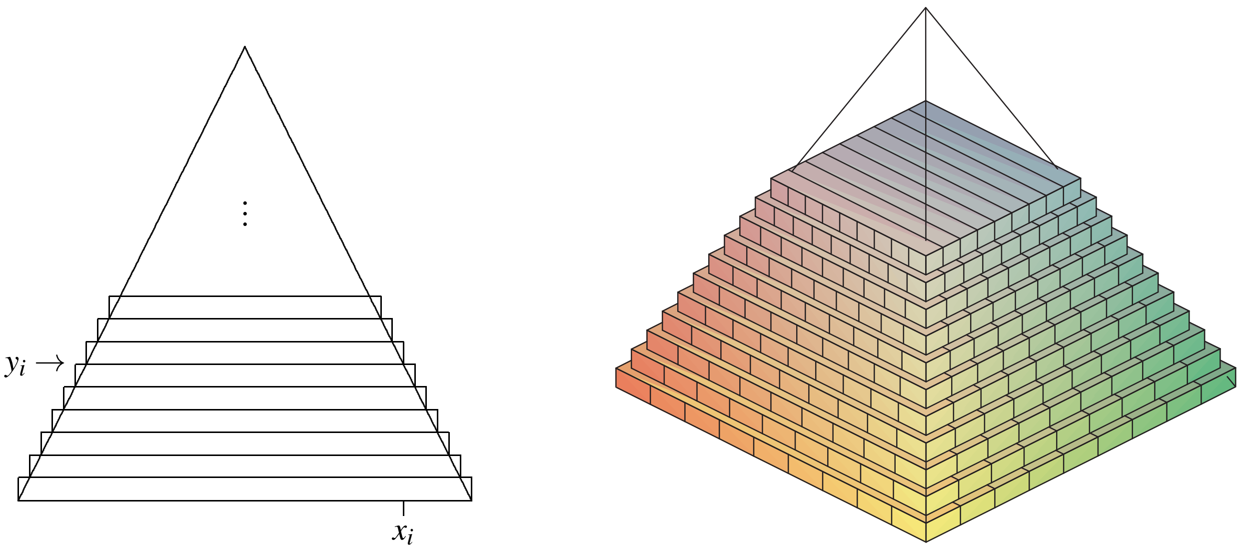


Figure 8.3: Volume of a pyramid approximated by rectangular prisms.

Example 8.7: Volume of a Pyramid

Find the volume of a pyramid with a square base that is 20 meters tall and 20 meters on a side at the base.

Solution. As with most of our applications of integration, we begin by asking how we might approximate the volume. Since we can easily compute the volume of a rectangular prism (that is, a "box"), we will use some boxes to approximate the volume of the pyramid, as shown in figure 8.3: On the left is a cross-sectional view, on the right is a 3D view of part of the pyramid with some of the boxes used to approximate the volume.

Each box has volume of the form $(2x_i)(2x_i)\Delta y$. Unfortunately, there are two variables here; fortunately, we can write x in terms of y: From the cross-sectional view we see that a height of 20 is achieved at the midpoint of the base. We will also position the cross-sectional view symmetrically about the y-axis. Thus at $x = 0$, $y = 20$, and we have a slope of $m = -2$. So

$$\begin{aligned} y &= -2x + b \\ 20 &= -2(0) + b \\ 20 &= b. \end{aligned}$$

Therefore, $y = 20 - 2x$, and in the terms of x: $x = 10 - y/2$ or $x_i = 10 - y_i/2$. Then the total volume is approximately

$$\sum_{i=0}^{n-1} 4(10 - y_i/2)^2 \Delta y$$

and in the limit we get the volume as the value of an integral:

$$\int_0^{20} 4(10 - y/2)^2\, dy = \int_0^{20} (20 - y)^2\, dy = -\left.\frac{(20-y)^3}{3}\right|_0^{20} = -\frac{0^3}{3} - -\frac{20^3}{3} = \frac{8000}{3}.$$

As you may know, the volume of a pyramid is $(1/3)(\text{height})(\text{area of base}) = (1/3)(20)(400)$, which agrees with our answer. ♣

Example 8.8: Volume of an Object

The base of a solid is the region between $f(x) = x^2 - 1$ and $g(x) = -x^2 + 1$, and its cross-sections perpendicular to the x-axis are equilateral triangles, as indicated in Figure 8.4. The solid has been truncated to show a triangular cross-section above $x = 1/2$. Find the volume of the solid.

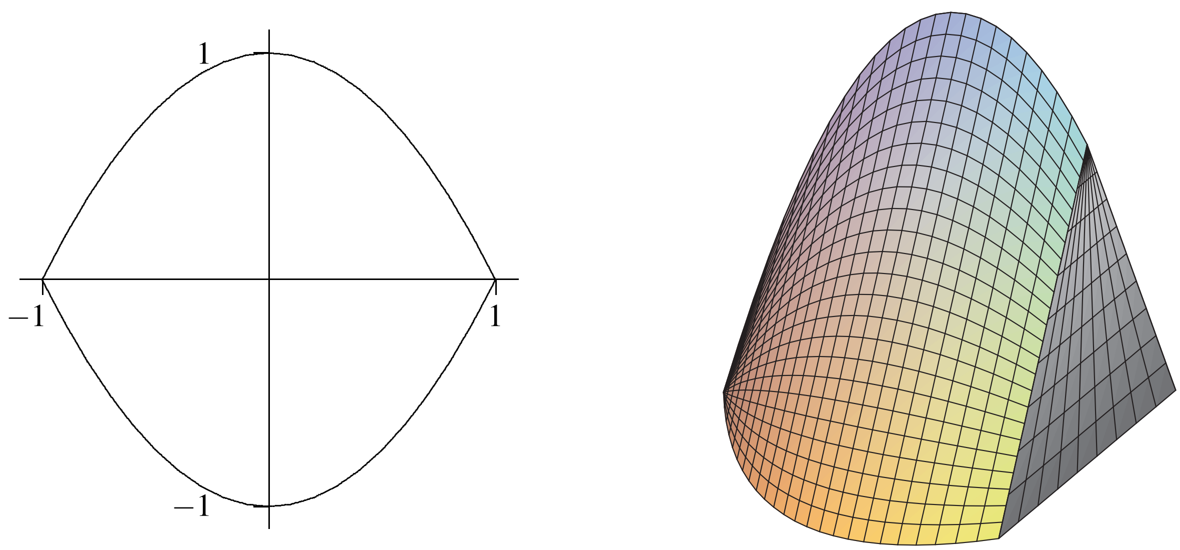


Figure 8.4: Solid with equilateral triangles as cross-sections.

Solution. A cross-section at a value x_i on the x-axis is a triangle with base $2(1 - x_i^2)$ and height $\sqrt{3}(1 - x_i^2)$, so the area of the cross-section is

$$\frac{1}{2}(\text{base})(\text{height}) = (1 - x_i^2)\sqrt{3}(1 - x_i^2),$$

and the volume of a thin "slab" is then

$$(1 - x_i^2)\sqrt{3}(1 - x_i^2)\Delta x.$$

Thus the total volume is

$$\int_{-1}^{1} \sqrt{3}(1 - x^2)^2\, dx = \frac{16}{15}\sqrt{3}.$$

♣

One easy way to get "nice" cross-sections is by rotating a plane figure around a line. For example, in Figure 8.5 we see a plane region under a curve and between two vertical lines; then the result of rotating this around the x-axis, and a typical circular cross-section is a circle.

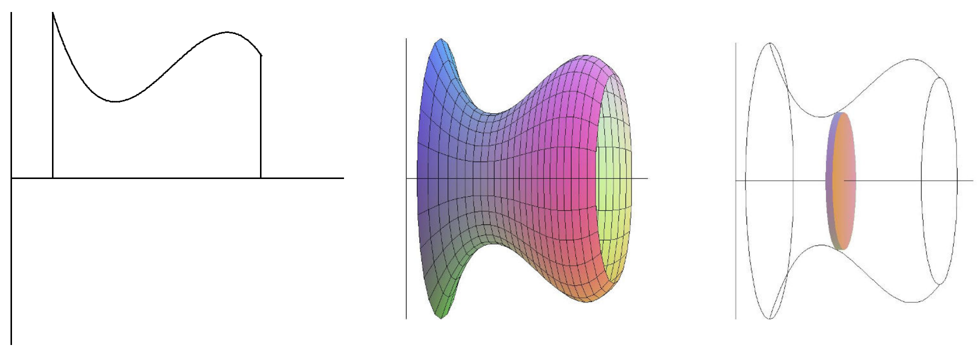


Figure 8.5: A solid of rotation.

Of course a real "slice" of this figure will not have straight sides, but we can approximate the volume of the slice by a cylinder or disk with circular top and bottom and straight sides; the volume of this disk will have the form $\pi r^2 \Delta x$. As long as we can write r in terms of x we can compute the volume by an integral.

Example 8.9: Volume of a Right Circular Cone

Find the volume of a right circular cone with base radius 10 and height 20. (A right circular cone is one with a circular base and with the tip of the cone directly over the center of the base.)

Solution. We can view this cone as produced by the rotation of the line $y = x/2$ rotated about the x-axis, as indicated in figure 8.6.

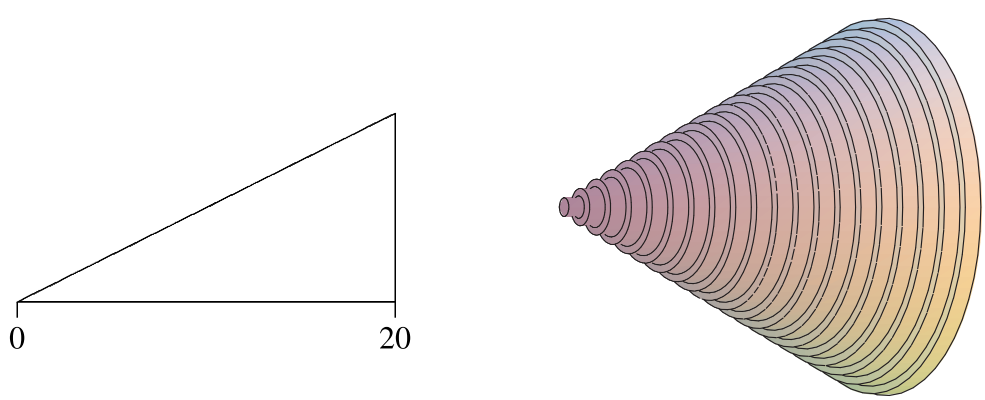


Figure 8.6: A region that generates a cone; approximating the volume by circular disks.

At a particular point on the x-axis, say x_i, the radius of the resulting cone is the y-coordinate of the corresponding point on the line, namely $y_i = x_i/2$. Thus the total volume is approximately

$$\sum_{i=0}^{n-1} \pi (x_i/2)^2 \, dx$$

and the exact volume is

$$\int_0^{20} \pi \frac{x^2}{4} \, dx = \frac{\pi}{4} \frac{20^3}{3} = \frac{2000\pi}{3}.$$

Note that we can instead do the calculation with a generic height and radius:

$$\int_0^h \pi \frac{r^2}{h^2} x^2 \, dx = \frac{\pi r^2}{h^2} \frac{h^3}{3} = \frac{\pi r^2 h}{3},$$

giving us the usual formula for the volume of a cone. ♣

Example 8.10: Volume of an Object with a Hole

Find the volume of the object generated when the area between $y = x^2$ and $y = x$ is rotated around the x-axis.

Solution. This solid has a "hole" in the middle; we can compute the volume by subtracting the volume of the hole from the volume enclosed by the outer surface of the solid. In figure 8.7 we show the region that is rotated, the resulting solid with the front half cut away, the cone that forms the outer surface, the horn-shaped hole, and a cross-section perpendicular to the x-axis.

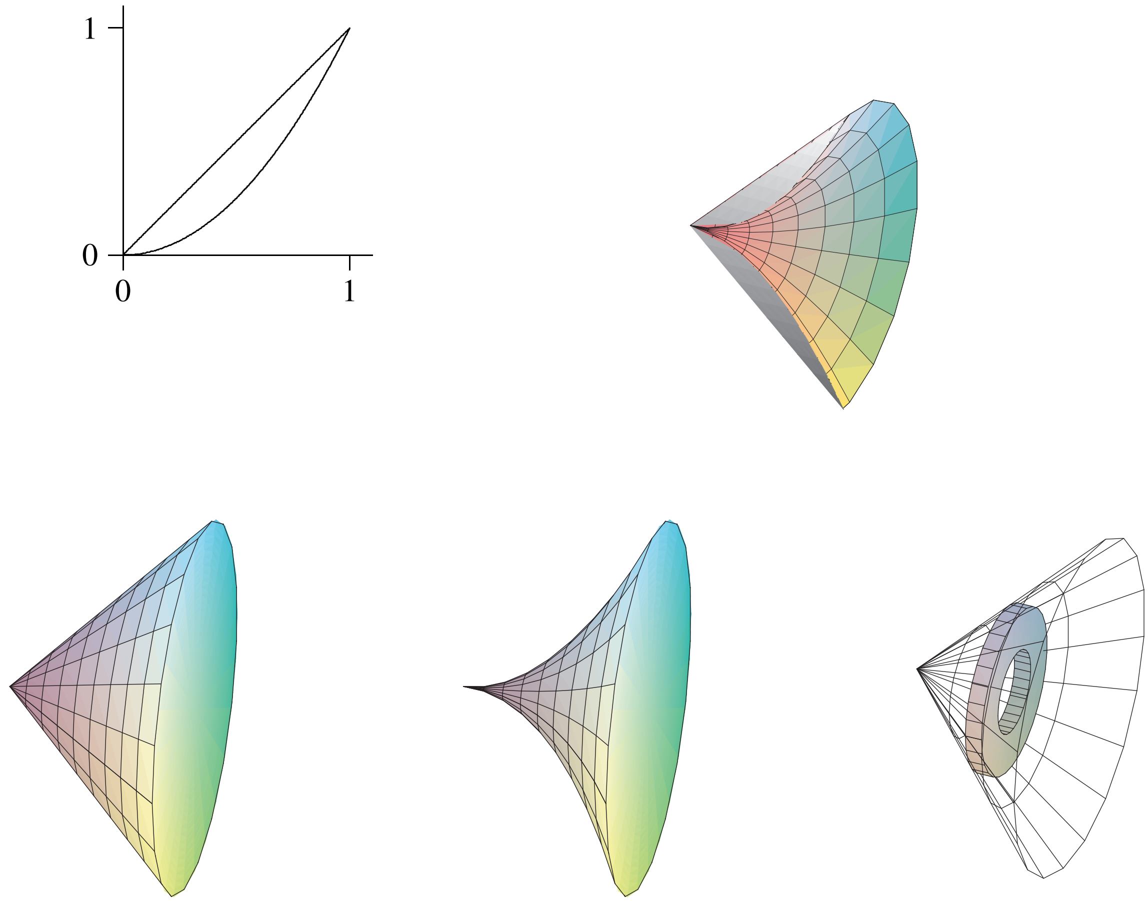


Figure 8.7: Solid with a hole, showing the outer cone and the shape to be removed to form the hole.

We have already computed the volume of a cone; in this case it is $\pi/3$. At a particular value of x, say x_i, the cross-section of the horn is a circle with radius x_i^2, so the volume of the horn is

$$\int_0^1 \pi(x^2)^2\,dx = \int_0^1 \pi x^4\,dx = \pi\frac{1}{5},$$

so the desired volume is $\pi/3 - \pi/5 = 2\pi/15$.

As with the area between curves, there is an alternate approach that computes the desired volume "all at once" by approximating the volume of the actual solid. We can approximate the volume of a slice of the solid with a washer-shaped volume, as indicated in Figure 8.7.

The volume of such a washer is the area of the face times the thickness. The thickness, as usual, is Δx, while the area of the face is the area of the outer circle minus the area of the inner circle, say $\pi R^2 - \pi r^2$, or $\pi(\text{TOP})^2 - \pi(\text{BOTTOM})^2$. In the present example, at a particular x_i, the radius R (The "TOP" function) is x_i and r (The "BOTTOM" function) x_i^2. Hence, the whole volume is

$$\int_0^1 \pi\left(\text{TOP}^2 - \text{BOTTOM}^2\right)dx = \int_0^1 \pi x^2 - \pi x^4\,dx = \pi\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Big|_0^1 = \pi\left(\frac{1}{3} - \frac{1}{5}\right) = \frac{2\pi}{15}.$$

Of course, what we have done here is exactly the same calculation as before, except we have in effect recomputed the volume of the outer cone. ♣

Suppose the region between $f(x) = x+1$ and $g(x) = (x-1)^2$ is rotated around the y-axis; see Figure 8.8. It is possible, but inconvenient, to compute the volume of the resulting solid by the method we have used so far. The problem is that there are two "kinds" of typical rectangles: Those that go from the line to the parabola and those that touch the parabola on both ends. To compute the volume using this approach, we need to break the problem into two parts and compute two integrals:

$$\pi \int_0^1 (1+\sqrt{y})^2 - (1-\sqrt{y})^2 \, dy + \pi \int_1^4 (1+\sqrt{y})^2 - (y-1)^2 \, dy = \frac{8}{3}\pi + \frac{65}{6}\pi = \frac{27}{2}\pi.$$

If instead we consider a typical vertical rectangle, but still rotate around the y-axis, we get a thin "shell" instead of a thin "washer". Note that "washers" are related to the area of a circle, πr^2, whereas "shells" are related to the surface area of a cylinder, $2\pi rh$. If we add up the volume of such thin shells we will get an approximation to the true volume. What is the volume of such a shell? Consider the shell at x_i. Imagine that we cut the shell vertically in one place and "unroll" it into a thin, flat sheet, namely the surface of a cylinder. This sheet will be almost a rectangular prism that is Δx thick, $f(x_i) - g(x_i)$ (TOP−BOTTOM) tall, and $2\pi x_i$ wide. The volume will then be approximately the volume of a rectangular prism with these dimensions: $2\pi x_i(f(x_i) - g(x_i))\Delta x$. If we add these up and take the limit as usual, we get the integral

$$\int_0^3 2\pi x(f(x) - g(x)) \, dx = \int_0^3 2\pi x \,(\text{TOP} - \text{BOTTOM}) \, dx = \int_0^3 2\pi x(x+1-(x-1)^2) \, dx = \frac{27}{2}\pi.$$

Not only does this accomplish the task with only one integral, the integral is somewhat easier than those in the previous calculation. Things are not always so neat, but it is often the case that one of the two methods will be simpler than the other, so it is worth considering both before starting to do calculations.

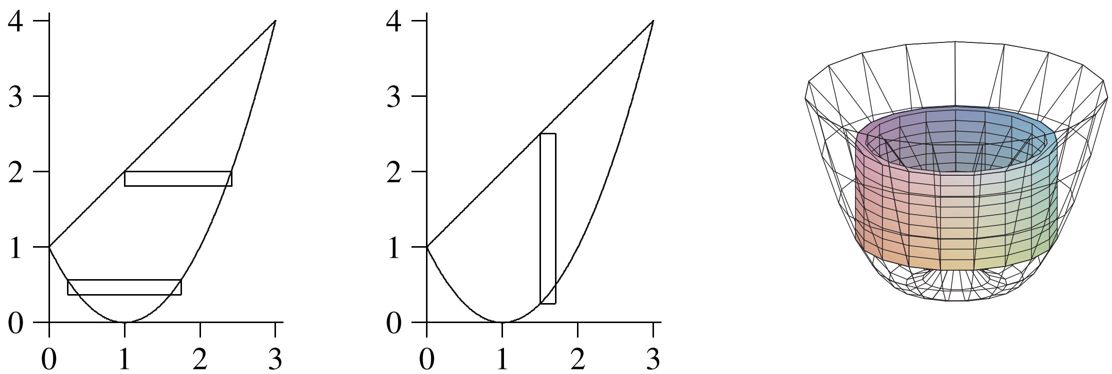


Figure 8.8: Computing volumes with "shells".

Example 8.11:

Suppose the area under $y = -x^2 + 1$ between $x = 0$ and $x = 1$ is rotated around the x-axis. Find the volume by both methods.

Solution. Using the disk method we obtain:

$$\int_0^1 \pi(1-x^2)^2 \, dx = \frac{8}{15}\pi.$$

Using the shell method we obtain:

$$\int_0^1 2\pi y\sqrt{1-y}\,dy = \frac{8}{15}\pi.$$

♣

Exercises for 8.3

Exercise 8.3.1 *Verify that* $\pi\int_0^1 (1+\sqrt{y})^2-(1-\sqrt{y})^2\,dy+\pi\int_1^4 (1+\sqrt{y})^2-(y-1)^2 = \frac{8}{3}\pi+\frac{65}{6}\pi = \frac{27}{2}\pi.$

Exercise 8.3.2 *Verify that* $\int_0^3 2\pi x(x+1-(x-1)^2)\,dx = \frac{27}{2}\pi.$

Exercise 8.3.3 *Verify that* $\int_0^1 \pi(1-x^2)^2\,dx = \frac{8}{15}\pi.$

Exercise 8.3.4 *Verify that* $\int_0^1 2\pi y\sqrt{1-y}\,dy = \frac{8}{15}\pi.$

Exercise 8.3.5 *Use integration to find the volume of the solid obtained by revolving the region bounded by $x+y=2$ and the x and y axes around the x-axis.*

Exercise 8.3.6 *Find the volume of the solid obtained by revolving the region bounded by $y=x-x^2$ and the x-axis around the x-axis.*

Exercise 8.3.7 *Find the volume of the solid obtained by revolving the region bounded by $y=\sqrt{\sin x}$ between $x=0$ and $x=\pi/2$, the y-axis, and the line $y=1$ around the x-axis.*

Exercise 8.3.8 *Let S be the region of the xy-plane bounded above by the curve $x^3y=64$, below by the line $y=1$, on the left by the line $x=2$, and on the right by the line $x=4$. Find the volume of the solid obtained by rotating S around:*

(a) the x-axis;

(b) the line $y=1$;

(c) the y-axis; and

(d) the line $x=2$.

Exercise 8.3.9 *The equation $x^2/9+y^2/4=1$ describes an ellipse. Find the volume of the solid obtained by rotating the ellipse around the x-axis and also around the y-axis. These solids are called* **ellipsoids**; *one is vaguely rugby-ball shaped, one is sort of flying-saucer shaped, or perhaps squished-beach-ball-shaped.*

Figure 8.9: Ellipsoids.

Exercise 8.3.10 *Use integration to compute the volume of a sphere of radius r. You should of course get the well-known formula* $4\pi r^3/3$*.*

Exercise 8.3.11 *A hemispheric bowl of radius r contains water to a depth h. Find the volume of water in the bowl.*

Exercise 8.3.12 *The base of a tetrahedron (a triangular pyramid) of height h is an equilateral triangle of side s. Its cross-sections perpendicular to an altitude are equilateral triangles. Express its volume V as an integral, and find a formula for V in terms of h and s. Verify that your answer is* $(1/3)(\textit{area of base})(\textit{height})$*.*

Exercise 8.3.13 *The base of a solid is the region between* $f(x) = \cos x$ *and* $g(x) = -\cos x$*,* $-\pi/2 \le x \le \pi/2$*, and its cross-sections perpendicular to the x-axis are squares. Find the volume of the solid.*

8.4 Average Value of a Function

The average of some finite set of values is a familiar concept. If, for example, the class scores on a quiz are 10, 9, 10, 8, 7, 5, 7, 6, 3, 2, 7, 8, then the average score is the sum of these numbers divided by the size of the class:

$$\text{average score} = \frac{10+9+10+8+7+5+7+6+3+2+7+8}{12} = \frac{82}{12} \approx 6.83.$$

Suppose that between $t = 0$ and $t = 1$ the speed of an object is $\sin(\pi t)$. What is the average speed of the object over that time? The question sounds as if it must make sense, yet we can't merely add up some number of speeds and divide, since the speed is changing continuously over the time interval.

To make sense of "average" in this context, we fall back on the idea of approximation. Consider the speed of the object at tenth of a second intervals: $\sin 0$, $\sin(0.1\pi)$, $\sin(0.2\pi)$, $\sin(0.3\pi)$,..., $\sin(0.9\pi)$. The average speed "should" be fairly close to the average of these ten speeds:

$$\frac{1}{10}\sum_{i=0}^{9} \sin(\pi i/10) \approx \frac{1}{10} 6.3 = 0.63.$$

Of course, if we compute more speeds at more times, the average of these speeds should be closer to the "real" average. If we take the average of n speeds at evenly spaced times, we get:

$$\frac{1}{n}\sum_{i=0}^{n-1} \sin(\pi i/n).$$

Here the individual times are $t_i = i/n$, so rewriting slightly we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \sin(\pi t_i).$$

This is almost the sort of sum that we know turns into an integral; what's apparently missing is Δt—but in fact, $\Delta t = 1/n$, the length of each subinterval. So rewriting again:

$$\sum_{i=0}^{n-1} \sin(\pi t_i)\frac{1}{n} = \sum_{i=0}^{n-1} \sin(\pi t_i)\Delta t.$$

Now this has exactly the right form, so that in the limit we get

$$\text{average speed} = \int_0^1 \sin(\pi t)\,dt = -\frac{\cos(\pi t)}{\pi}\bigg|_0^1 = -\frac{\cos(\pi)}{\pi} + \frac{\cos(0)}{\pi} = \frac{2}{\pi} \approx 0.6366 \approx 0.64.$$

It's not entirely obvious from this one simple example how to compute such an average in general. Let's look at a somewhat more complicated case. Suppose that the velocity of an object is $16t^2 + 5$ feet per second. What is the average velocity between $t = 1$ and $t = 3$? Again we set up an approximation to the average:

$$\frac{1}{n}\sum_{i=0}^{n-1} 16t_i^2 + 5,$$

where the values t_i are evenly spaced times between 1 and 3. Once again we are "missing" Δt, and this time $1/n$ is not the correct value. What is Δt in general? It is the length of a subinterval; in this case we take the interval $[1,3]$ and divide it into n subintervals, so each has length $(3-1)/n = 2/n = \Delta t$. Now with the usual "multiply and divide by the same thing" trick we can rewrite the sum:

$$\frac{1}{n}\sum_{i=0}^{n-1} 16t_i^2 + 5 = \frac{1}{3-1}\sum_{i=0}^{n-1} (16t_i^2 + 5)\frac{3-1}{n} = \frac{1}{2}\sum_{i=0}^{n-1} (16t_i^2 + 5)\frac{2}{n} = \frac{1}{2}\sum_{i=0}^{n-1} (16t_i^2 + 5)\Delta t.$$

In the limit this becomes

$$\frac{1}{2}\int_1^3 16t^2 + 5\,dt = \frac{1}{2}\frac{446}{3} = \frac{223}{3}.$$

Does this seem reasonable? Let's picture it: In Figure 8.10 we see the velocity function together with the horizontal line $y = 223/3 \approx 74.3$. Certainly the height of the horizontal line looks at least plausible for the average height of the curve.

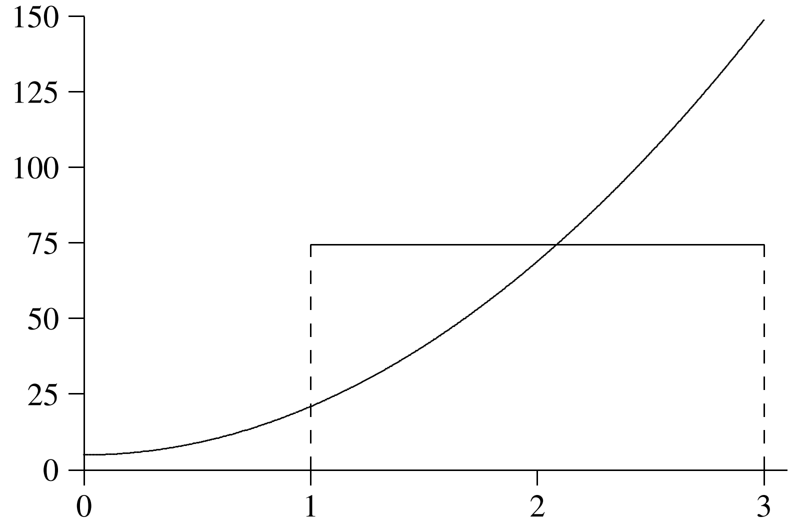


Figure 8.10: Average velocity.

Here's another way to interpret "average" that may make our computation appear even more reasonable. The object of our example goes a certain distance between $t = 1$ and $t = 3$. If instead the object were to travel at the average speed over the same time, it should go the same distance. At an average speed of $223/3$ feet per second for two seconds the object would go $446/3$ feet. How far does it actually go? We know how to compute this:

$$\int_1^3 v(t)\,dt = \int_1^3 16t^2 + 5\,dt = \frac{446}{3}.$$

So now we see that another interpretation of the calculation is:

$$\frac{1}{2}\int_1^3 16t^2 + 5\,dt = \frac{1}{2}\frac{446}{3} = \frac{223}{3}$$

which is the total distance traveled divided by the time in transit, namely, the usual interpretation of average speed.

In the case of speed, or more properly velocity, we can always interpret "average" as total (net) distance divided by time. However, in the case of a different sort of quantity this interpretation does not obviously apply, while the approximation approach always does. We might interpret the same problem geometrically: What is the average height of $16x^2 + 5$ on the interval $[1,3]$? We approximate this in exactly the same way, by adding up many sample heights and dividing by the number of samples. In the limit we get the same result:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} 16x_i^2 + 5 = \frac{1}{2}\int_1^3 16x^2 + 5\,dx = \frac{1}{2}\frac{446}{3} = \frac{223}{3}.$$

We can interpret this result in a slightly different way. The area under $y = 16x^2 + 5$ above $[1,3]$ is

$$\int_1^3 16t^2 + 5\,dt = \frac{446}{3}.$$

The area under $y = 223/3$ over the same interval $[1,3]$ is simply the area of a rectangle that is 2 by $223/3$ with area $446/3$. So the average height of a function is the height of the horizontal line that produces the same area over the given interval.

Exercises for 8.4

Exercise 8.4.1 *Find the average height of* $\cos x$ *over the intervals* $[0, \pi/2]$, $[-\pi/2, \pi/2]$, *and* $[0, 2\pi]$.

Exercise 8.4.2 *Find the average height of* x^2 *over the interval* $[-2, 2]$.

Exercise 8.4.3 *Find the average height of* $1/x^2$ *over the interval* $[1, A]$.

Exercise 8.4.4 *Find the average height of* $\sqrt{1-x^2}$ *over the interval* $[-1, 1]$.

Exercise 8.4.5 *An object moves with velocity* $v(t) = -t^2 + 1$ *feet per second between* $t = 0$ *and* $t = 2$. *Find the average velocity and the average speed of the object between* $t = 0$ *and* $t = 2$.

Exercise 8.4.6 *The observation deck on the 102nd floor of the Empire State Building is 1,224 feet above the ground. If a steel ball is dropped from the observation deck its velocity at time* t *is approximately* $v(t) = -32t$ *feet per second. Find the average speed between the time it is dropped and the time it hits the ground, and find its speed when it hits the ground.*

8.5 Work

A fundamental concept in classical physics is **work**: If an object is moved in a straight line against a force F for a distance d the work done is $W = Fd$.

Example 8.12: Constant Force

How much work is done in lifting a 10 pound weight vertically a distance of 5 feet?

Solution. The force due to gravity on a 10 pound weight is 10 pounds at the surface of the earth, and it does not change appreciably over 5 feet. The work done is $W = 10 \cdot 5 = 50$ foot-pounds. ♣

In reality few situations are so simple. The force might not be constant over the range of motion, as in the next example.

Example 8.13: Lifting a Weight

How much work is done in lifting a 10 pound weight from the surface of the earth to an orbit 100 miles above the surface?

Solution. Over 100 miles the force due to gravity does change significantly, so we need to take this into account. The force exerted on a 10 pound weight at a distance r from the center of the earth is $F = k/r^2$ and by definition it is 10 when r is the radius of the earth (we assume the earth is a sphere). How can we approximate the work done? We divide the path from the surface to orbit into n small subpaths. On each subpath the force due to gravity is roughly constant, with value k/r_i^2 at distance r_i. The work to raise the object from r_i to r_{i+1} is thus approximately $k/r_i^2\Delta r$ and the total work is approximately

$$\sum_{i=0}^{n-1} \frac{k}{r_i^2}\Delta r,$$

or in the limit

$$W = \int_{r_0}^{r_1} \frac{k}{r^2}\,dr,$$

where r_0 is the radius of the earth and r_1 is r_0 plus 100 miles. The work is

$$W = \int_{r_0}^{r_1} \frac{k}{r^2}\,dr = -\left.\frac{k}{r}\right|_{r_0}^{r_1} = -\frac{k}{r_1} + \frac{k}{r_0}.$$

Using $r_0 = 20925525$ feet we have $r_1 = 21453525$. The force on the 10 pound weight at the surface of the earth is 10 pounds, so $10 = k/20925525^2$, giving $k = 4378775965256250$. Then

$$-\frac{k}{r_1} + \frac{k}{r_0} = \frac{491052320000}{95349} \approx 5150052 \quad \text{foot-pounds.}$$

Note that if we assume the force due to gravity is 10 pounds over the whole distance we would calculate the work as $10(r_1 - r_0) = 10 \cdot 100 \cdot 5280 = 5280000$, somewhat higher since we don't account for the weakening of the gravitational force. ♣

Example 8.14: Lifting an Object

How much work is done in lifting a 10 kilogram object from the surface of the earth to a distance D from the center of the earth?

Solution. This is the same problem as before in different units, and we are not specifying a value for D. As before

$$W = \int_{r_0}^{D} \frac{k}{r^2}\,dr = -\left.\frac{k}{r}\right|_{r_0}^{D} = -\frac{k}{D} + \frac{k}{r_0}.$$

While "weight in pounds" is a measure of force, "weight in kilograms" is a measure of mass. To convert to force we need to use Newton's law $F = ma$. At the surface of the earth the acceleration due to gravity is approximately 9.8 meters per second squared, so the force is $F = 10 \cdot 9.8 = 98$. The units here are "kilogram-meters per second squared" or "kg m/s^2", also known as a Newton (N), so $F = 98$ N. The radius of the earth is approximately 6378.1 kilometers or 6378100 meters. Now the problem proceeds as before. From $F = k/r^2$ we compute k: $98 = k/6378100^2$, $k = 3.986655642 \cdot 10^{15}$. Then the work is:

$$W = -\frac{k}{D} + 6.250538000 \cdot 10^8 \quad \text{Newton-meters.}$$

As D increases W of course gets larger, since the quantity being subtracted, $-k/D$, gets smaller. But note that the work W will never exceed $6.250538000 \cdot 10^8$, and in fact will approach this value as D gets larger. In short, with a finite amount of work, namely $6.250538000 \cdot 10^8$ N-m, we can lift the 10 kilogram object as far as we wish from earth. ♣

Next is an example in which the force is constant, but there are many objects moving different distances.

Example 8.15: Multiple Objects Moving

Suppose that a water tank is shaped like a right circular cone with the tip at the bottom, and has height 10 meters and radius 2 meters at the top. If the tank is full, how much work is required to pump all the water out over the top?

Solution. Here we have a large number of atoms of water that must be lifted different distances to get to the top of the tank. Fortunately, we don't really have to deal with individual atoms—we can consider all the atoms at a given depth together.

To approximate the work, we can divide the water in the tank into horizontal sections, approximate the volume of water in a section by a thin disk, and compute the amount of work required to lift each disk to the top of the tank. As usual, we take the limit as the sections get thinner and thinner to get the total work.

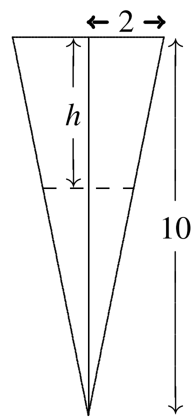


Figure 8.11: Cross-section of a conical water tank.

At depth h the circular cross-section through the tank has radius $r = (10-h)/5$, by similar triangles, and area $\pi(10-h)^2/25$. A section of the tank at depth h thus has volume approximately $\pi(10-h)^2/25\Delta h$ and so contains $\sigma\pi(10-h)^2/25\Delta h$ kilograms of water, where σ is the density of water in kilograms per cubic meter; $\sigma \approx 1000$. The force due to gravity on this much water is $9.8\sigma\pi(10-h)^2/25\Delta h$, and finally, this section of water must be lifted a distance h, which requires $h9.8\sigma\pi(10-h)^2/25\Delta h$ Newton-meters of work. The total work is therefore

$$W = \frac{9.8\sigma\pi}{25}\int_0^{10} h(10-h)^2\,dh = \frac{980000}{3}\pi \approx 1026254 \quad \text{Newton-meters.}$$

♣

A spring has a "natural length," its length if nothing is stretching or compressing it. If the spring is either stretched or compressed the spring provides an opposing force; according to **Hooke's Law** the magnitude of this force is proportional to the distance the spring has been stretched or compressed: $F = kx$. The constant of proportionality, k, of course depends on the spring. Note that x here represents the *change* in length from the natural length.

Example 8.16: Compressing a Spring

Suppose $k = 5$ for a given spring that has a natural length of 0.1 meters. Suppose a force is applied that compresses the spring to length 0.08. What is the magnitude of the force?

Solution. Assuming that the constant k has appropriate dimensions (namely, kg/s^2), the force is $5(0.1-0.08) = 5(0.02) = 0.1$ Newtons. ♣

Example 8.17: Compressing a Spring (continued)

How much work is done in compressing the spring in the previous example from its natural length to 0.08 meters? From 0.08 meters to 0.05 meters? How much work is done to stretch the spring from 0.1 meters to 0.15 meters?

Solution. We can approximate the work by dividing the distance that the spring is compressed (or stretched) into small subintervals. Then the force exerted by the spring is approximately constant over the subinterval, so the work required to compress the spring from x_i to x_{i+1} is approximately $5(x_i - 0.1)\Delta x$. The total work is approximately

$$\sum_{i=0}^{n-1} 5(x_i - 0.1)\Delta x$$

and in the limit

$$W = \int_{0.1}^{0.08} 5(x-0.1)\,dx = \left.\frac{5(x-0.1)^2}{2}\right|_{0.1}^{0.08} = \frac{5(0.08-0.1)^2}{2} - \frac{5(0.1-0.1)^2}{2} = \frac{1}{1000}\text{ N-m.}$$

The other values we seek simply use different limits. To compress the spring from 0.08 meters to 0.05 meters takes

$$W = \int_{0.08}^{0.05} 5(x-0.1)\,dx = \left.\frac{5(x-0.1)^2}{2}\right|_{0.08}^{0.05} = \frac{5(0.05-0.1)^2}{2} - \frac{5(0.08-0.1)^2}{2} = \frac{21}{4000}\quad\text{N-m}$$

and to stretch the spring from 0.1 meters to 0.15 meters requires

$$W = \int_{0.1}^{0.15} 5(x-0.1)\,dx = \left.\frac{5(x-0.1)^2}{2}\right|_{0.1}^{0.15} = \frac{5(0.15-0.1)^2}{2} - \frac{5(0.1-0.1)^2}{2} = \frac{1}{160}\quad\text{N-m.}$$

♣

Exercises for 8.5

Exercise 8.5.1 *How much work is done in lifting a 100 kilogram weight from the surface of the earth to an orbit 35,786 kilometers above the surface of the earth?*

Exercise 8.5.2 *How much work is done in lifting a 100 kilogram weight from an orbit 1000 kilometers above the surface of the earth to an orbit 35,786 kilometers above the surface of the earth?*

Exercise 8.5.3 *A water tank has the shape of an upright cylinder with radius $r = 1$ meter and height 10 meters. If the depth of the water is 5 meters, how much work is required to pump all the water out the top of the tank?*

Exercise 8.5.4 *Suppose the tank of the previous problem is lying on its side, so that the circular ends are vertical, and that it has the same amount of water as before. How much work is required to pump the water out the top of the tank (which is now 2 meters above the bottom of the tank)?*

Exercise 8.5.5 *A water tank has the shape of the bottom half of a sphere with radius $r = 1$ meter. If the tank is full, how much work is required to pump all the water out the top of the tank?*

Exercise 8.5.6 *A spring has constant $k = 10$ kg/s^2. How much work is done in compressing it $1/10$ meter from its natural length?*

Exercise 8.5.7 *A force of 2 Newtons will compress a spring from 1 meter (its natural length) to 0.8 meters. How much work is required to stretch the spring from 1.1 meters to 1.5 meters?*

Exercise 8.5.8 *A 20 meter long steel cable has density 2 kilograms per meter, and is hanging straight down. How much work is required to lift the entire cable to the height of its top end?*

Exercise 8.5.9 *The cable in the previous problem has a 100 kilogram bucket of concrete attached to its lower end. How much work is required to lift the entire cable and bucket to the height of its top end?*

Exercise 8.5.10 *Consider again the cable and bucket of the previous problem. How much work is required to lift the bucket 10 meters by raising the cable 10 meters? (The top half of the cable ends up at the height of the top end of the cable, while the bottom half of the cable is lifted 10 meters.)*

8.6 Center of Mass

Suppose a beam is 10 meters long, and that there are three weights on the beam: a 10 kilogram weight 3 meters from the left end, a 5 kilogram weight 6 meters from the left end, and a 4 kilogram weight 8 meters from the left end. Where should a fulcrum be placed so that the beam balances? Let's assign a scale to the beam, from 0 at the left end to 10 at the right, so that we can denote locations on the beam simply as x coordinates; the weights are at $x = 3$, $x = 6$, and $x = 8$, as in Figure 8.12.

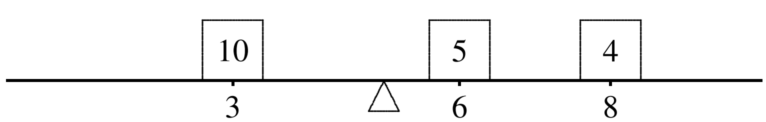


Figure 8.12: A beam with three masses.

Suppose to begin with that the fulcrum is placed at $x = 5$. What will happen? Each weight applies a force to the beam that tends to rotate it around the fulcrum; this effect is measured by a quantity called **torque**, proportional to the mass times the distance from the fulcrum. Of course, weights on different sides of the fulcrum rotate the beam in opposite directions. We can distinguish this by using a signed distance in the formula for torque. So with the fulcrum at 5, the torques induced by the three weights will be proportional to $(3-5)10 = -20$, $(6-5)5 = 5$, and $(8-5)4 = 12$. For the beam to balance, the sum of the torques must be zero; since the sum is $-20+5+12 = -3$, the beam rotates counter-clockwise, and to get the beam to balance we need to move the fulcrum to the left. To calculate exactly where the fulcrum should be, we let $\bar{x}$ denote the location of the fulcrum when the beam is in balance. The total torque on the beam is then $(3-\bar{x})10+(6-\bar{x})5+(8-\bar{x})4 = 92-19\bar{x}$. Since the beam balances at $\bar{x}$ it must be that $92-19\bar{x} = 0$ or $\bar{x} = 92/19 \approx 4.84$, that is, the fulcrum should be placed at $x = 92/19$ to balance the beam.

Now suppose that we have a beam with varying density—some portions of the beam contain more mass than other portions of the same size. We want to figure out where to put the fulcrum so that the beam balances.

m_0	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9

Figure 8.13: A solid beam.

Example 8.18: Balance Point of a Beam

Find the balance point of a solid beam, illustrated in Figure 8.13, assuming the beam is 10 meters long and that the density is $1+x$ kilograms per meter at location x on the beam.

Solution. To approximate the solution, we can think of the beam as a sequence of weights "on" a beam. For example, we can think of the portion of the beam between $x = 0$ and $x = 1$ as a weight sitting at $x = 0$, the portion between $x = 1$ and $x = 2$ as a weight sitting at $x = 1$, and so on, as indicated in Figure 8.13. We then approximate the mass of the weights by assuming that each portion of the beam has constant density. So the mass of the first weight is approximately $m_0 = (1+0)1 = 1$ kilograms, namely, $(1+0)$ kilograms per meter times 1 meter. The second weight is $m_1 = (1+1)1 = 2$ kilograms, and so on to the tenth weight with $m_9 = (1+9)1 = 10$ kilograms. So in this case the total torque is

$$(0-\bar{x})m_0+(1-\bar{x})m_1+\cdots+(9-\bar{x})m_9 = (0-\bar{x})1+(1-\bar{x})2+\cdots+(9-\bar{x})10.$$

If we set this to zero and solve for $\bar{x}$ we get $\bar{x} = 6$. In general, if we divide the beam into n portions, the mass of weight number i will be $m_i = (1+x_i)(x_{i+1}-x_i) = (1+x_i)\Delta x$ and the torque induced by weight number i will be $(x_i-\bar{x})m_i = (x_i-\bar{x})(1+x_i)\Delta x$. The total torque is then

$$\begin{aligned}
&(x_0-\bar{x})(1+x_0)\Delta x+(x_1-\bar{x})(1+x_1)\Delta x+\cdots+(x_{n-1}-\bar{x})(1+x_{n-1})\Delta x \\
&\qquad = \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x - \sum_{i=0}^{n-1} \bar{x}(1+x_i)\Delta x \\
&\qquad = \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x - \bar{x}\sum_{i=0}^{n-1}(1+x_i)\Delta x.
\end{aligned}$$

If we set this equal to zero and solve for $\bar{x}$ we get an approximation to the balance point of the beam:

$$\begin{aligned}
0 &= \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x - \bar{x}\sum_{i=0}^{n-1}(1+x_i)\Delta x \\
\bar{x}\sum_{i=0}^{n-1}(1+x_i)\Delta x &= \sum_{i=0}^{n-1} x_i(1+x_i)\Delta x \\
\bar{x} &= \frac{\sum_{i=0}^{n-1} x_i(1+x_i)\Delta x}{\sum_{i=0}^{n-1}(1+x_i)\Delta x}.
\end{aligned}$$

The denominator of this fraction has a very familiar interpretation. Consider one term of the sum in the denominator: $(1+x_i)\Delta x$. This is the density near x_i times a short length, Δx, which in other words is approximately the mass of the beam between x_i and x_{i+1}. When we add these up we get approximately the mass of the beam.

Now each of the sums in the fraction has the right form to turn into an integral, which in turn gives us the exact value of $\bar{x}$:

$$\bar{x} = \frac{\int_0^{10} x(1+x)\,dx}{\int_0^{10}(1+x)\,dx}.$$

The numerator of this fraction is called the **moment** of the system around zero:

$$\int_0^{10} x(1+x)\,dx = \int_0^{10} x+x^2\,dx = \frac{1150}{3},$$

and the denominator is the mass of the beam:

$$\int_0^{10}(1+x)\,dx = 60,$$

and the balance point, officially called the **center of mass** is

$$\bar{x} = \frac{1150}{3}\frac{1}{60} = \frac{115}{18} \approx 6.39.$$

♣

It should be apparent that there was nothing special about the density function $\sigma(x) = 1+x$ or the length of the beam, or even that the left end of the beam is at the origin. In general, if the density of the beam is $\sigma(x)$ and the beam covers the interval $[a,b]$, the moment of the beam around zero is

$$M_0 = \int_a^b x\sigma(x)\,dx$$

and the total mass of the beam is

$$M = \int_a^b \sigma(x)\,dx$$

and the center of mass is at

$$\bar{x} = \frac{M_0}{M}.$$

Example 8.19: Center of Mass of a Beam

Suppose a beam lies on the x-axis between 20 and 30, and has density function $\sigma(x) = x - 19$. Find the center of mass.

Solution. This is the same as the previous example except that the beam has been moved. Note that the density at the left end is $20 - 19 = 1$ and at the right end is $30 - 19 = 11$, as before. Hence the center of mass must be at approximately $20 + 6.39 = 26.39$. Let's see how the calculation works out.

$$\begin{aligned}
M_0 &= \int_{20}^{30} x(x-19)\,dx = \int_{20}^{30} x^2 - 19x\,dx = \left.\frac{x^3}{3} - \frac{19x^2}{2}\right|_{20}^{30} = \frac{4750}{3} \\
M &= \int_{20}^{30} x - 19\,dx = \left.\frac{x^2}{2} - 19x\right|_{20}^{30} = 60 \\
\frac{M_0}{M} &= \frac{4750}{3}\frac{1}{60} = \frac{475}{18} \approx 26.39.
\end{aligned}$$

♣

Example 8.20: Centroid of a Flat Plate

Suppose a flat plate of uniform density has the shape contained by $y = x^2$, $y = 1$, and $x = 0$, in the first quadrant. Find the center of mass. (Since the density is constant, the center of mass depends only on the shape of the plate, not the density, or in other words, this is a purely geometric quantity. In such a case the center of mass is called the **centroid***.)*

Solution. This is a two dimensional problem, but it can be solved as if it were two one dimensional problems: we need to find the x and y coordinates of the center of mass, $\bar{x}$ and $\bar{y}$, and fortunately we can do these independently. Imagine looking at the plate edge on, from below the x-axis. The plate will appear to be a beam, and the mass of a short section of the "beam", say between x_i and x_{i+1}, is the mass of a strip of the plate between x_i and x_{i+1}. See Figure 8.14 showing the plate from above and as it appears edge on.

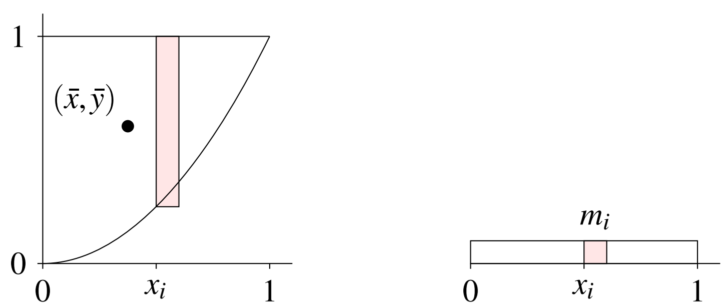


Figure 8.14: Center of mass for a two dimensional plate.

Since the plate has uniform density we may as well assume that $\sigma = 1$. Then the mass of the plate between x_i and x_{i+1} is approximately $m_i = \sigma(1-x_i^2)\Delta x = (1-x_i^2)\Delta x$. Now we can compute the moment around the y-axis:

$$M_y = \int_0^1 x(1-x^2)\,dx = \frac{1}{4}$$

and the total mass

$$M = \int_0^1 (1-x^2)\,dx = \frac{2}{3}$$

and finally

$$\bar{x} = \frac{1}{4}\frac{3}{2} = \frac{3}{8}.$$

Next we do the same thing to find $\bar{y}$. The mass of the plate between y_i and y_{i+1} is approximately $n_i = \sqrt{y}\Delta y$, so

$$M_x = \int_0^1 y\sqrt{y}\,dy = \frac{2}{5}$$

and

$$\bar{y} = \frac{2}{5}\frac{3}{2} = \frac{3}{5},$$

since the total mass M is the same. The center of mass is shown in Figure 8.14. ♣

Example 8.21: Center of Mass under Cosine

Find the center of mass of a thin, uniform plate whose shape is the region between $y = \cos x$ and the x-axis between $x = -\pi/2$ and $x = \pi/2$.

Solution. It is clear that $\bar{x} = 0$, but for practice let's compute it anyway. We will need the total mass, so we compute it first:

$$M = \int_{-\pi/2}^{\pi/2} \cos x\,dx = \sin x\Big|_{-\pi/2}^{\pi/2} = 2.$$

The moment around the y-axis is

$$M_y = \int_{-\pi/2}^{\pi/2} x\cos x\,dx = \cos x + x\sin x\Big|_{-\pi/2}^{\pi/2} = 0$$

and the moment around the x-axis is

$$M_x = \int_0^1 y\cdot 2\arccos y\,dy = y^2\arccos y - \frac{y\sqrt{1-y^2}}{2} + \frac{\arcsin y}{2}\Bigg|_0^1 = \frac{\pi}{4}.$$

Thus

$$\bar{x} = \frac{0}{2}, \quad \bar{y} = \frac{\pi}{8} \approx 0.393.$$

♣

Exercises for 8.6

Exercise 8.6.1 *A beam 10 meters long has density* $\sigma(x) = x^2$ *at distance* x *from the left end of the beam. Find the center of mass* $\bar{x}$.

Exercise 8.6.2 *A beam 10 meters long has density* $\sigma(x) = \sin(\pi x/10)$ *at distance* x *from the left end of the beam. Find the center of mass* $\bar{x}$.

Exercise 8.6.3 *A beam 4 meters long has density* $\sigma(x) = x^3$ *at distance* x *from the left end of the beam. Find the center of mass* $\bar{x}$.

Exercise 8.6.4 *Verify that* $\displaystyle\int 2x \arccos x \, dx = x^2 \arccos x - \frac{x\sqrt{1-x^2}}{2} + \frac{\arcsin x}{2} + C.$

Exercise 8.6.5 *A thin plate lies in the region between* $y = x^2$ *and the x-axis between* $x = 1$ *and* $x = 2$. *Find the centroid.*

Exercise 8.6.6 *A thin plate fills the upper half of the unit circle* $x^2 + y^2 = 1$. *Find the centroid.*

Exercise 8.6.7 *A thin plate lies in the region contained by* $y = x$ *and* $y = x^2$. *Find the centroid.*

Exercise 8.6.8 *A thin plate lies in the region contained by* $y = 4 - x^2$ *and the x-axis. Find the centroid.*

Exercise 8.6.9 *A thin plate lies in the region contained by* $y = x^{1/3}$ *and the x-axis between* $x = 0$ *and* $x = 1$. *Find the centroid.*

Exercise 8.6.10 *A thin plate lies in the region contained by* $\sqrt{x} + \sqrt{y} = 1$ *and the axes in the first quadrant. Find the centroid.*

Exercise 8.6.11 *A thin plate lies in the region between the circle* $x^2 + y^2 = 4$ *and the circle* $x^2 + y^2 = 1$, *above the x-axis. Find the centroid.*

Exercise 8.6.12 *A thin plate lies in the region between the circle* $x^2 + y^2 = 4$ *and the circle* $x^2 + y^2 = 1$ *in the first quadrant. Find the centroid.*

Exercise 8.6.13 *A thin plate lies in the region between the circle* $x^2 + y^2 = 25$ *and the circle* $x^2 + y^2 = 16$ *above the x-axis. Find the centroid.*

8.7 Arc Length

Here is another geometric application of the integral: Find the length of a portion of a curve. As usual, we need to think about how we might approximate the length, and turn the approximation into an integral.

We already know how to compute one simple arc length, that of a line segment. If the endpoints are $P_0(x_0,y_0)$ and $P_1(x_1,y_1)$ then the length of the segment is the distance between the points, $\sqrt{(x_1-x_0)^2+(y_1-y_0)^2}$, from the Pythagorean theorem, as illustrated in Figure 8.15.

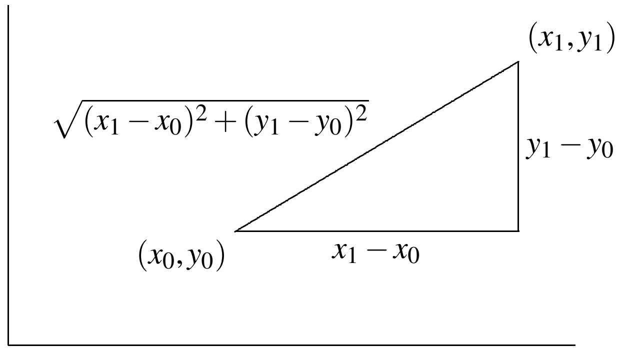


Figure 8.15: The length of a line segment.

Now if the graph of f is "nice" (say, differentiable) it appears that we can approximate the length of a portion of the curve with line segments, and that as the number of segments increases, and their lengths decrease, the sum of the lengths of the line segments will approach the true arc length; see Figure 8.16.

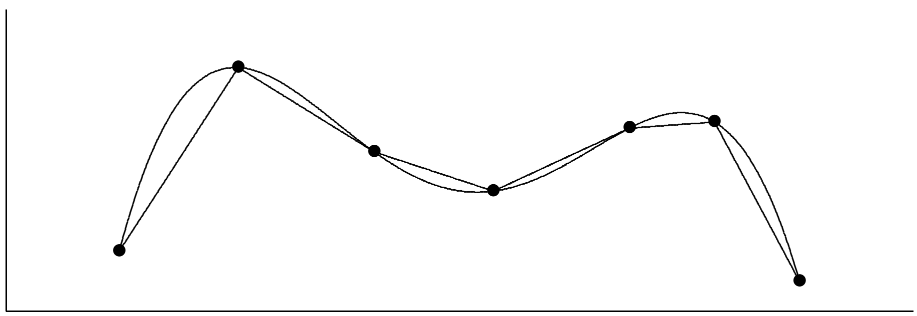


Figure 8.16: Approximating arc length with line segments.

Now we need to write a formula for the sum of the lengths of the line segments, in a form that we know becomes an integral in the limit. So we suppose we have divided the interval $[a,b]$ into n subintervals as usual, each with length $\Delta x=(b-a)/n$, and endpoints $a=x_0$, x_1, x_2, …, $x_n=b$. The length of a typical line segment, joining $(x_i,f(x_i))$ to $(x_{i+1},f(x_{i+1}))$, is $\sqrt{(\Delta x)^2+(f(x_{i+1})-f(x_i))^2}$. By the Mean Value Theorem, there is a number t_i in (x_i,x_{i+1}) such that $f'(t_i)\Delta x=f(x_{i+1})-f(x_i)$, so the length of the line segment can be written as

$$\sqrt{(\Delta x)^2+(f'(t_i))^2\Delta x^2}=\sqrt{1+(f'(t_i))^2}\,\Delta x.$$

Then arc length is:

$$\lim_{n\to\infty}\sum_{i=0}^{n-1}\sqrt{1+(f'(t_i))^2}\,\Delta x=\int_a^b\sqrt{1+(f'(x))^2}\,dx.$$

Note that the sum looks a bit different than others we have encountered, because the approximation contains a t_i instead of an x_i. In the past we have always used left endpoints (namely, x_i) to get a representative value of f on $[x_i,x_{i+1}]$; now we are using a different point, but the principle is the same.

To summarize, to compute the length of a curve on the interval $[a,b]$, we compute the integral

$$\int_a^b \sqrt{1+(f'(x))^2}\,dx.$$

Unfortunately, integrals of this form are typically difficult or impossible to compute exactly, because usually none of our methods for finding antiderivatives will work. In practice this means that the integral will usually have to be approximated.

Example 8.22: Circumference of a Circle

Let $f(x)=\sqrt{r^2-x^2}$, the upper half circle of radius r. The length of this curve is half the circumference, namely πr. Compute this with the arc length formula.

Solution. The derivative f' is $-x/\sqrt{r^2-x^2}$ so the integral is

$$\int_{-r}^{r}\sqrt{1+\frac{x^2}{r^2-x^2}}\,dx=\int_{-r}^{r}\sqrt{\frac{r^2}{r^2-x^2}}\,dx=r\int_{-r}^{r}\sqrt{\frac{1}{r^2-x^2}}\,dx.$$

Using a trigonometric substitution, we find the antiderivative, namely $\arcsin(x/r)$. Notice that the integral is improper at both endpoints, as the function $\sqrt{1/(r^2-x^2)}$ is undefined when $x=\pm r$. So we need to compute

$$\lim_{D\to -r^+}\int_D^0\sqrt{\frac{1}{r^2-x^2}}\,dx+\lim_{D\to r^-}\int_0^D\sqrt{\frac{1}{r^2-x^2}}\,dx.$$

This is not difficult, and has value π, so the original integral, with the extra r in front, has value πr as expected. ♣

Exercises for 8.7

Exercise 8.7.1 *Find the arc length of $f(x)=x^{3/2}$ on $[0,2]$.*

Exercise 8.7.2 *Find the arc length of $f(x)=x^2/8-\ln x$ on $[1,2]$.*

Exercise 8.7.3

Find the arc length of $f(x)=(1/3)(x^2+2)^{3/2}$ on the interval $[0,a]$.

Exercise 8.7.4 *Find the arc length of $f(x)=\ln(\sin x)$ on the interval $[\pi/4,\pi/3]$.*

Exercise 8.7.5 *Let $a>0$. Show that the length of $y=\cosh x$ on $[0,a]$ is equal to $\int_0^a \cosh x\,dx$.*

Exercise 8.7.6 *Find the arc length of $f(x)=\cosh x$ on $[0,\ln 2]$.*

Exercise 8.7.7 *Set up the integral to find the arc length of* $\sin x$ *on the interval* $[0,\pi]$*; do not evaluate the integral. If you have access to appropriate software, approximate the value of the integral.*

Exercise 8.7.8 *Set up the integral to find the arc length of* $y = xe^{-x}$ *on the interval* $[2,3]$*; do not evaluate the integral. If you have access to appropriate software, approximate the value of the integral.*

Exercise 8.7.9 *Find the arc length of* $y = e^x$ *on the interval* $[0,1]$. *(This can be done exactly; it is a bit tricky and a bit long.)*

8.8 Surface Area

Another geometric question that arises naturally is: "What is the surface area of a volume?" For example, what is the surface area of a sphere? More advanced techniques are required to approach this question in general, but we can compute the areas of some volumes generated by revolution.

As usual, the question is: How might we approximate the surface area? For a surface obtained by rotating a curve around an axis, we can take a polygonal approximation to the curve, as in the last section, and rotate it around the same axis. This gives a surface composed of many "truncated cones"; a truncated cone is called a **frustum** of a cone. Figure 8.17 illustrates this approximation.

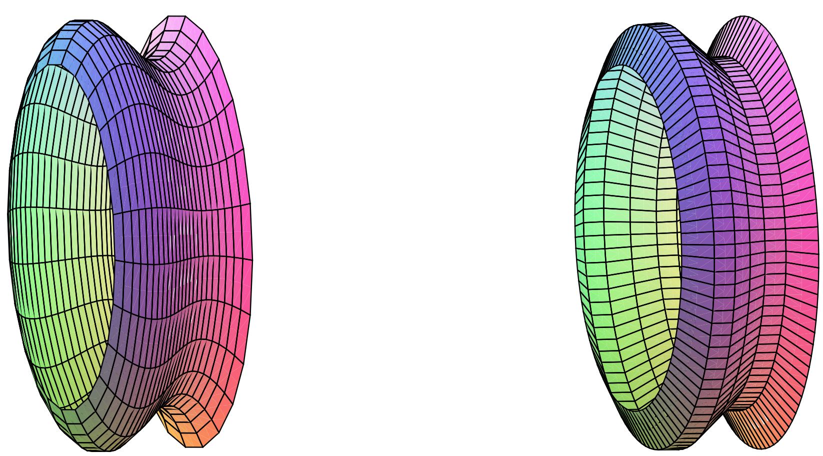


Figure 8.17: Approximating a surface (left) by portions of cones (right).

So we need to be able to compute the area of a frustum of a cone. Since the frustum can be formed by removing a small cone from the top of a larger one, we can compute the desired area if we know the surface area of a cone. Suppose a right circular cone has base radius r and slant height h. If we cut the cone from the vertex to the base circle and flatten it out, we obtain a sector of a circle with radius h and arc length $2\pi r$, as in Figure 8.18. The angle at the center, in radians, is then $2\pi r/h$, and the area of the cone is equal to the area of the sector of the circle. Let A be the area of the sector; since the area of the entire circle is πh^2, we have

$$\frac{A}{\pi h^2} = \frac{2\pi r/h}{2\pi}$$
$$A = \pi r h.$$

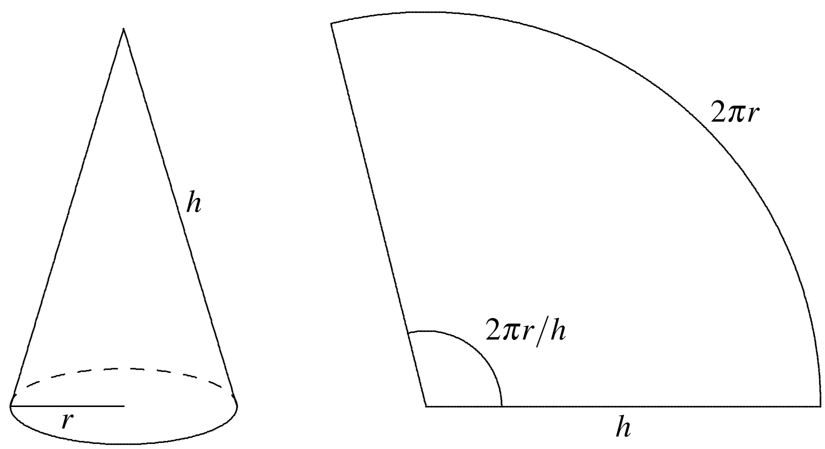


Figure 8.18: The area of a cone.

Now suppose we have a frustum of a cone with slant height h and radii r_0 and r_1, as in Figure 8.19. The area of the entire cone is $\pi r_1(h_0 + h)$, and the area of the small cone is $\pi r_0 h_0$; thus, the area of the frustum is $\pi r_1(h_0 + h) - \pi r_0 h_0 = \pi((r_1 - r_0)h_0 + r_1 h)$. By similar triangles,

$$\frac{h_0}{r_0} = \frac{h_0 + h}{r_1}.$$

With a bit of algebra this becomes $(r_1 - r_0)h_0 = r_0 h$; substitution into the area gives

$$\pi((r_1 - r_0)h_0 + r_1 h) = \pi(r_0 h + r_1 h) = \pi h(r_0 + r_1) = 2\pi \frac{r_0 + r_1}{2} h = 2\pi r h.$$

The final form is particularly easy to remember, with r equal to the average of r_0 and r_1, as it is also the formula for the area of a cylinder. (Think of a cylinder of radius r and height h as the frustum of a cone of infinite height.)

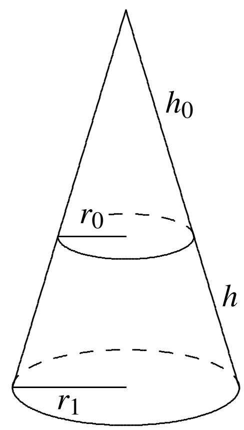


Figure 8.19: The area of a frustum.

Now we are ready to approximate the area of a surface of revolution. On one subinterval, the situation is as shown in Figure 8.20. When the line joining two points on the curve is rotated around the x-axis, it forms a frustum of a cone. The area is

$$2\pi rh = 2\pi\frac{f(x_i)+f(x_{i+1})}{2}\sqrt{1+(f'(t_i))^2}\,\Delta x.$$

Here $\sqrt{1+(f'(t_i))^2}\,\Delta x$ is the length of the line segment, as we found in the previous section. Assuming f is a continuous function, there must be some x_i^* in $[x_i, x_{i+1}]$ such that $(f(x_i)+f(x_{i+1}))/2 = f(x_i^*)$, so the approximation for the surface area is

$$\sum_{i=0}^{n-1} 2\pi f(x_i^*)\sqrt{1+(f'(t_i))^2}\,\Delta x.$$

This is not quite the sort of sum we have seen before, as it contains two different values in the interval $[x_i, x_{i+1}]$, namely x_i^* and t_i. Nevertheless, using more advanced techniques than we have available here, it turns out that

$$\lim_{n\to\infty}\sum_{i=0}^{n-1} 2\pi f(x_i^*)\sqrt{1+(f'(t_i))^2}\,\Delta x = \int_a^b 2\pi f(x)\sqrt{1+(f'(x))^2}\,dx$$

is the surface area we seek. (Roughly speaking, this is because while x_i^* and t_i are distinct values in $[x_i, x_{i+1}]$, they get closer and closer to each other as the length of the interval shrinks.)

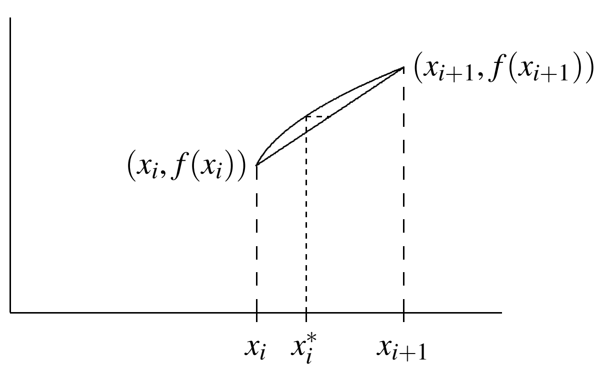


Figure 8.20: One subinterval.

Example 8.23: Surface Area of a Sphere

Compute the surface area of a sphere of radius r.

Solution. The sphere can be obtained by rotating the graph of $f(x) = \sqrt{r^2 - x^2}$ about the x-axis. The derivative f' is $-x/\sqrt{r^2-x^2}$, so the surface area is given by

$$\begin{aligned} A &= 2\pi \int_{-r}^{r} \sqrt{r^2-x^2}\sqrt{1+\frac{x^2}{r^2-x^2}}\,dx \\ &= 2\pi \int_{-r}^{r} \sqrt{r^2-x^2}\sqrt{\frac{r^2}{r^2-x^2}}\,dx \\ &= 2\pi \int_{-r}^{r} r\,dx = 2\pi r \int_{-r}^{r} 1\,dx = 4\pi r^2 \end{aligned}$$

♣

If the curve is rotated around the y axis, the formula is nearly identical, because the length of the line segment we use to approximate a portion of the curve doesn't change. Instead of the radius $f(x_i^*)$, we use the new radius $\bar{x}_i = (x_i + x_{i+1})/2$, and the surface area integral becomes

$$\int_a^b 2\pi x \sqrt{1+(f'(x))^2}\,dx.$$

Example 8.24: Surface Around y-axis

Compute the area of the surface formed when $f(x) = x^2$ between 0 and 2 is rotated around the y-axis.

Solution. We compute $f'(x) = 2x$, and then

$$2\pi \int_0^2 x\sqrt{1+4x^2}\,dx = \frac{\pi}{6}(17^{3/2}-1),$$

by a simple substitution. ♣

Exercises for 8.8

Exercise 8.8.1 *Compute the area of the surface formed when $f(x) = 2\sqrt{1-x}$ between -1 and 0 is rotated around the x-axis.*

Exercise 8.8.2 *Compute the surface area of example 8.24 by rotating $f(x) = \sqrt{x}$ around the x-axis.*

Exercise 8.8.3 *Compute the area of the surface formed when $f(x) = x^3$ between 1 and 3 is rotated around the x-axis.*

Exercise 8.8.4 *Compute the area of the surface formed when* $f(x) = 2 + \cosh(x)$ *between* 0 *and* 1 *is rotated around the x-axis.*

Exercise 8.8.5 *Consider the surface obtained by rotating the graph of* $f(x) = 1/x$, $x \geq 1$, *around the x-axis. This surface is called* **Gabriel's horn** *or* **Toricelli's trumpet***. Show that Gabriel's horn has infinite surface area.*

Exercise 8.8.6 *Consider the circle* $(x-2)^2 + y^2 = 1$. *Sketch the surface obtained by rotating this circle about the y-axis. (The surface is called a* **torus***.) What is the surface area?*

Exercise 8.8.7 *Consider the ellipse with equation* $x^2/4 + y^2 = 1$. *If the ellipse is rotated around the x-axis it forms an* **ellipsoid***. Compute the surface area.*

Exercise 8.8.8 *Generalize the preceding result: rotate the ellipse given by* $x^2/a^2 + y^2/b^2 = 1$ *about the x-axis and find the surface area of the resulting ellipsoid. You should consider two cases, when* $a > b$ *and when* $a < b$. *Compare to the area of a sphere.*

9. Sequences and Series

Consider the following sum:

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots+\frac{1}{2^i}+\cdots$$

The dots at the end indicate that the sum goes on forever. Does this make sense? Can we assign a numerical value to an infinite sum? While at first it may seem difficult or impossible, we have certainly done something similar when we talked about one quantity getting "closer and closer" to a fixed quantity. Here we could ask whether, as we add more and more terms, the sum gets closer and closer to some fixed value. That is, look at

$$\begin{aligned}
\frac{1}{2} &= \frac{1}{2}\\
\frac{3}{4} &= \frac{1}{2}+\frac{1}{4}\\
\frac{7}{8} &= \frac{1}{2}+\frac{1}{4}+\frac{1}{8}\\
\frac{15}{16} &= \frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}
\end{aligned}$$

and so on, and consider whether these values have a limit. It seems likely that they do, namely 1. In fact, as we will see, it's not hard to show that

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots+\frac{1}{2^i}=\frac{2^i-1}{2^i}=1-\frac{1}{2^i}$$

and then

$$\lim_{i\to\infty} 1-\frac{1}{2^i},$$

which gets closer and closer to 1 as i gets larger.

There is a context in which we already implicitly accept this notion of infinite sum without really thinking of it as a sum: The representation of a real number as an infinite decimal. For example,

$$0.3333\bar{3}=\frac{3}{10}+\frac{3}{100}+\frac{3}{1000}+\frac{3}{10000}+\cdots=\frac{1}{3},$$

or likewise

$$3.14159\ldots=3+\frac{1}{10}+\frac{4}{100}+\frac{1}{1000}+\frac{5}{10000}+\frac{9}{100000}+\cdots=\pi.$$

An infinite sum is called a **series**, and is usually written using the same sigma notation that we encountered in Chapter 6. In this case, however, we use ∞ to indicate that there is no 'last term'. The series we first examined can be written as

$$\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^i}=\sum_{i=1}^{\infty}\frac{1}{2^i}$$

A related notion that will aid our investigations is that of a **sequence**. A sequence is just an ordered (possibly infinite) list of numbers. For example,

$$1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$$

We will begin by learning some useful facts about sequences.

9.1 Sequences

While the idea of a sequence of numbers, $a_1, a_2, a_3, \ldots$ is straightforward, it is useful to think of a sequence as a function. We have dealt with functions whose domains are the real numbers, or a subset of the real numbers, like $f(x) = \sin x$. A sequence can be regarded as a function with domain as the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ or the non-negative integers, $\mathbb{Z}^{\geq 0} = \{0, 1, 2, 3, \ldots\}$. The range of the function is still allowed to be the set of all real numbers; we say that a sequence is a function $f \colon \mathbb{N} \to \mathbb{R}$. Sequences are commonly denoted in several different, but equally acceptable ways:

$$a_1, a_2, a_3, \ldots$$
$$\{a_n\}_{n=1}^{\infty}$$
$$\{f(n)\}_{n=1}^{\infty}$$

As with functions of the real numbers, we will most often encounter sequences that can be expressed by a formula. We have already seen the sequence $a_i = f(i) = 1 - 1/2^i$. Some other simple examples are:

$$f(i) = \frac{i}{i+1}$$
$$f(n) = \frac{1}{2^n}$$
$$f(n) = \sin(n\pi/6)$$
$$f(i) = \frac{(i-1)(i+2)}{2^i}$$

Frequently these formulas will make sense if thought of either as functions with domain $\mathbb{R}$ or $\mathbb{N}$, though occasionally one will make sense for integer values only.

The main question of interest when dealing with sequences is what happens to the terms as we go further and further down the list. In particular, as i becomes extremely large, does a_i get closer to one specific value? This is reminiscent of a question we asked in Chapter 3, when looking at limits of functions. In fact, the problems are closely related and we define the limit of a sequence in a way similar to Definition 3.16.

Definition 9.1: Limit of a Sequence

Suppose that $\{a_n\}_{n=1}^{\infty}$ is a sequence. We say that $\lim_{n\to\infty} a_n = L$ if for every $\varepsilon > 0$ there is an $N > 0$ so that whenever $n > N$, $|a_n - L| < \varepsilon$. If $\lim_{n\to\infty} a_n = L$ we say that the sequence **converges** *to L, otherwise it* **diverges**.

Intuitively, $\lim_{n\to\infty} a_2 = L$ means that the further we go in the sequence, the closer the terms get to L.

Example 9.2: Exponential Sequence

Show that $\{2^{1/n}\}_{n=1}^{\infty}$ *converges to 1.*

Solution. Suppose $\varepsilon > 0$. Then let $N = \frac{1}{\log_2(1+\varepsilon)}$. Note that $N > 0$. Now if $n > N$, then

$$\begin{aligned} n &> \frac{1}{\log_2(1+\varepsilon)} \\ \log_2(1+\varepsilon) &> \frac{1}{n} \\ 1+\varepsilon &> 2^{1/n} \\ \varepsilon &> 2^{1/n} - 1 \\ \varepsilon &> \left|2^{1/n} - 1\right| \end{aligned}$$

♣

Note that, as in Chapter 3, we generally need to work "backwards" from the last line of the proof to determine how to choose N. Having done so, we write the actual proof as we have done here to show that this value of N 'works'.

If a sequence is defined by a formula $\{f(i)\}_{i=1}^{\infty}$, we can often expand the domain of the function f to the set of all (or almost all) real numbers. For example, $f(i) = \frac{1}{i}$ is defined for all non-zero real numbers.

When this happens, we can sometimes find the limit of the sequence $\{f(i)\}_{i=1}^{\infty}$ more easily by finding the limit of the function $f(x)$, $x \in \mathbb{R}$, as x approaches infinity.

Theorem 9.3: Limit of a Sequence

If $\lim_{x\to\infty} f(x) = L$, *where* $f : \mathbb{R} \to \mathbb{R}$, *then* $\{f(i)\}_{i=1}^{\infty}$ *converges to* L.

Proof. This follows immediately from Definitions 3.16 and 9.1. ♣

Hereafter we will use the convention that x refers to a real-valued variable and i and n are integer-valued.

Example 9.4: Sequence of $1/n$

Show that $\{\frac{1}{n}\}_{n=0}^{\infty}$ *converges to 0.*

Solution. Since $\lim_{x\to\infty} \frac{1}{x} = 0$, $\lim_{n\to\infty} \frac{1}{n} = 0$. ♣

Note that the converse of Theorem 9.3 is not true.

Let $f(n) = \sin(n\pi)$. This is the sequence

$$\sin(0\pi), \sin(1\pi), \sin(2\pi), \sin(3\pi), \ldots = 0, 0, 0, 0, \ldots$$

since $\sin(n\pi) = 0$ when n is an integer. Thus $\lim\limits_{n\to\infty} f(n) = 0$. But $\lim\limits_{x\to\infty} f(x)$, when x is real, does not exist: as x gets bigger and bigger, the values $\sin(x\pi)$ do not get closer and closer to a single value, but take on all values between -1 and 1 over and over. In general, whenever you want to know $\lim\limits_{n\to\infty} f(n)$ you should first attempt to compute $\lim\limits_{x\to\infty} f(x)$, since if the latter exists it is also equal to the first limit. But if for some reason $\lim\limits_{x\to\infty} f(x)$ does not exist, it may still be true that $\lim\limits_{n\to\infty} f(n)$ exists, but you'll have to figure out another way to compute it.

It is occasionally useful to think of the graph of a sequence. Since the function is defined only for integer values, the graph is just a sequence of points. In Figure 9.1 we see the graphs of two sequences and the graphs of the corresponding real functions.

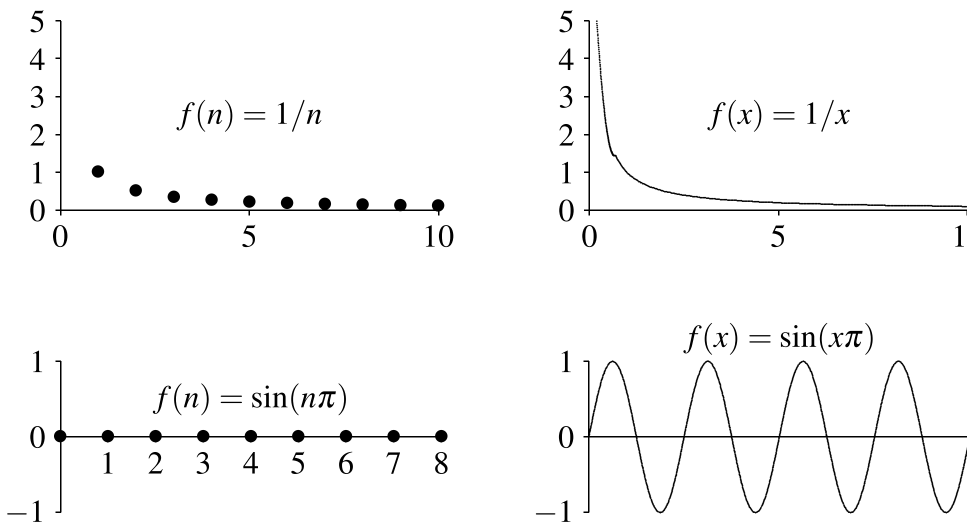


Figure 9.1: Graphs of sequences and their corresponding real functions.

Not surprisingly, the properties of limits of real functions translate into properties of sequences quite easily. Theorem 3.8 about limits becomes:

Theorem 9.5: Properties of Sequences

Suppose that $\lim\limits_{n\to\infty} a_n = L$ *and* $\lim\limits_{n\to\infty} b_n = M$ *and* k *is some constant. Then*

$$\lim_{n\to\infty} ka_n = k\lim_{n\to\infty} a_n = kL$$
$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n = L + M$$
$$\lim_{n\to\infty} (a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n = L - M$$
$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n = LM$$
$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n} = \frac{L}{M}, \text{ if } M \text{ is not } 0$$

Likewise the Squeeze Theorem (3.37) becomes:

Theorem 9.6: Squeeze Theorem for Sequences

Suppose that $a_n \le b_n \le c_n$ *for all* $n > N$, *for some* N. *If* $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$, *then* $\lim_{n\to\infty} b_n = L$.

And a final useful fact:

Theorem 9.7: Absolute Value Sequence

$\lim_{n\to\infty} |a_n| = 0$ *if and only if* $\lim_{n\to\infty} a_n = 0$.

This says simply that the size of $|a_n|$ gets close to zero if and only if a_n gets close to zero.

Example 9.8: Convergence of a Rational Fraction

Determine whether $\left\{\frac{n}{n+1}\right\}_{n=0}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Solution. Defining $f(x) = \frac{x}{x+1}$ we obtain

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

Thus the sequence converges to 1. ♣

Example 9.9: Convergence of Ratio with Natural Logarithm

Determine whether $\left\{\frac{\ln n}{n}\right\}_{n=1}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Solution. We compute

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

using L'Hôpital's Rule. Thus the sequence converges to 0. ♣

Example 9.10: Alternating Ones

Determine whether $\{(-1)^n\}_{n=0}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Solution. $f(x) = (-1)^x$ is undefined for irrational values of x so $\lim_{x\to\infty}(-1)^x$ does not exist. However, the sequence has a very simple pattern:

$$1, -1, 1, -1, 1 \ldots$$

and clearly diverges. ♣

Example 9.11: Convergence of Exponential

Determine whether $\{(-1/2)^n\}_{n=0}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Solution. We consider the sequence $\{|(-1/2)^n|\}_{n=0}^{\infty} = \{(1/2)^n\}_{n=0}^{\infty}$. Then

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

so by Theorem 9.7 the sequence converges to 0. ♣

Example 9.12: Using the Squeeze Theorem for Sequences

Determine whether $\{(\sin n)/\sqrt{n}\}_{n=1}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Solution. Since $|\sin n| \le 1$, $0 \le |\sin n/\sqrt{n}| \le 1/\sqrt{n}$ and we can use Theorem 9.6 with $a_n = 0$ and $c_n = 1/\sqrt{n}$. Since $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} c_n = 0$, $\lim\limits_{n\to\infty} \sin n/\sqrt{n} = 0$ and the sequence converges to 0. ♣

Example 9.13: Geometric Sequence

Let r *be a fixed real number. Determine when* $\{r^n\}_{n=0}^{\infty}$ *coverges.*

Solution. A particularly common and useful sequence is $\{r^n\}_{n=0}^{\infty}$, for various values of r. Some are quite easy to understand: If $r = 1$ the sequence converges to 1 since every term is 1, and likewise if $r = 0$ the sequence converges to 0. If $r = -1$ this is the sequence of Example 9.10 and diverges. If $r > 1$ or $r < -1$ the terms r^n get large without limit, so the sequence diverges. If $0 < r < 1$ then the sequence converges to 0. If $-1 < r < 0$ then $|r^n| = |r|^n$ and $0 < |r| < 1$, so the sequence $\{|r|^n\}_{n=0}^{\infty}$ converges to 0, so also $\{r^n\}_{n=0}^{\infty}$ converges to 0. converges. In summary, $\{r^n\}$ converges precisely when $-1 < r \le 1$ in which case

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & \text{if } -1 < r < 1 \\ 1 & \text{if } r = 1 \end{cases}$$

♣

Sequences of this form, or the more general form $\{kr^n\}_{n=0}^{\infty}$, are called **geometric sequences** or **geometric progressions**. They are encountered in a large variety of mathematical and real-world applications.

Sometimes we will not be able to determine the limit of a sequence, but we still would like to know whether it converges. In some cases we can determine this even without being able to compute the limit.

A sequence is called **increasing** or sometimes **strictly increasing** if $a_i < a_{i+1}$ for all i. It is called **non-decreasing** or sometimes (unfortunately) **increasing** if $a_i \le a_{i+1}$ for all i. Similarly a sequence is **decreasing** if $a_i > a_{i+1}$ for all i and **non-increasing** if $a_i \ge a_{i+1}$ for all i. If a sequence has any of these properties it is called **monotonic**.

Example 9.14:

The sequence

$$\left\{\frac{2^i-1}{2^i}\right\}_{i=1}^{\infty} = \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots,$$

is increasing, and

$$\left\{\frac{n+1}{n}\right\}_{i=1}^{\infty} = \frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \ldots$$

is decreasing.

A sequence is **bounded above** if there is some number N such that $a_n \leq N$ for every n, and **bounded below** if there is some number N such that $a_n \geq N$ for every n. If a sequence is bounded above and bounded below it is **bounded**. If a sequence $\{a_n\}_{n=0}^{\infty}$ is increasing or non-decreasing it is bounded below (by a_0), and if it is decreasing or non-increasing it is bounded above (by a_0). Finally, with all this new terminology we can state an important theorem.

Theorem 9.15: Bounded Monotonic Sequence

If a sequence is bounded and monotonic then it converges.

We will not prove this, but the proof appears in many calculus books. It is not hard to believe: suppose that a sequence is increasing and bounded, so each term is larger than the one before, yet never larger than some fixed value N. The terms must then get closer and closer to some value between a_0 and N. It need not be N, since N may be a "too-generous" upper bound; the limit will be the smallest number that is above all of the terms a_i.

Example 9.16:

Determine whether $\left\{\frac{2^i-1}{2^i}\right\}_{i=1}^{\infty}$ *converges.*

Solution. For every $i \geq 1$ we have $0 < (2^i-1)/2^i < 1$, so the sequence is bounded, and we have already observed that it is necessary. Therefore, the sequence converges. ♣

We don't actually need to know that a sequence is monotonic to apply this theorem—it is enough to know that the sequence is "eventually" monotonic, that is, that at some point it becomes increasing or decreasing. For example, the sequence 10, 9, 8, 15, 3, 21, 4, $3/4$, $7/8$, $15/16$, $31/32, \ldots$ is not increasing, because among the first few terms it is not. But starting with the term $3/4$ it is increasing, so the theorem tells us that the sequence $3/4, 7/8, 15/16, 31/32, \ldots$ converges. Since convergence depends only on what happens as n gets large, adding a few terms at the beginning can't turn a convergent sequence into a divergent one.

Example 9.17:

Show that $\{n^{1/n}\}$ *converges.*

Solution. We first show that this sequence is decreasing, that is, that $n^{1/n} > (n+1)^{1/(n+1)}$. Consider the real function $f(x) = x^{1/x}$ when $x \geq 1$. We can compute the derivative, $f'(x) = x^{1/x}(1-\ln x)/x^2$, and note that when $x \geq 3$ this is negative. Since the function has negative slope, $n^{1/n} > (n+1)^{1/(n+1)}$ when $n \geq 3$. Since all terms of the sequence are positive, the sequence is decreasing and bounded when $n \geq 3$, and so the sequence converges. (As it happens, we can compute the limit in this case, but we know it converges even without knowing the limit; see Exercise 9.1.1.) ♣

Example 9.18:

Show that $\{n!/n^n\}$ converges.

Solution. If we look at the ratio of successive terms we see that:

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}}\frac{n^n}{n!} = \frac{(n+1)!}{n!}\frac{n^n}{(n+1)^{n+1}} = \frac{n+1}{n+1}\left(\frac{n}{n+1}\right)^n = \left(\frac{n}{n+1}\right)^n < 1.$$

Therefore $a_{n+1} < a_n$, and so the sequence is decreasing. Since all terms are positive, it is also bounded, and so it must converge. (Again it is possible to compute the limit; see Exercise 9.1.2.) ♣

Exercises for 9.1

Exercise 9.1.1 *Compute* $\lim\limits_{x\to\infty} x^{1/x}$.

Exercise 9.1.2 *Use the squeeze theorem to show that* $\lim\limits_{n\to\infty} \dfrac{n!}{n^n} = 0$.

Exercise 9.1.3 *Determine whether* $\{\sqrt{n+47} - \sqrt{n}\}_{n=0}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Exercise 9.1.4 *Determine whether* $\left\{\dfrac{n^2+1}{(n+1)^2}\right\}_{n=0}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Exercise 9.1.5 *Determine whether* $\left\{\dfrac{n+47}{\sqrt{n^2+3n}}\right\}_{n=1}^{\infty}$ *converges or diverges. If it converges, compute the limit.*

Exercise 9.1.6 *Determine whether* $\left\{\dfrac{2^n}{n!}\right\}_{n=0}^{\infty}$ *converges or diverges.*

9.2 Series

While much more can be said about sequences, we now turn to our principal interest, series. Recall that a series, roughly speaking, is the sum of a sequence: If $\{a_n\}_{n=0}^{\infty}$ is a sequence then the associated series is

$$\sum_{i=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots$$

Associated with a series is a second sequence, called the **s**equence of partial sums $\{s_n\}_{n=0}^{\infty}$:

$$s_n = \sum_{i=0}^{n} a_i.$$

So

$$s_0 = a_0, \quad s_1 = a_0 + a_1, \quad s_2 = a_0 + a_1 + a_2, \quad \ldots$$

A series converges if the sequence of partial sums converges, and otherwise the series diverges.

If $\{kx^n\}_{n=0}^{\infty}$ is a geometric sequence, then the associated series $\sum_{i=0}^{\infty} kx^i$ is called a geometric series.

Theorem 9.19: Geometric Series Convergence

If $|x| < 1$, the geometric series $\sum_i kx^i$ converges to $\frac{k}{1-x}$, otherwise the series diverges (unless $k = 0$).

Proof. If $a_n = kx^n$, $\sum_{n=0}^{\infty} a_n$ is called a **geometric series**. A typical partial sum is

$$s_n = k + kx + kx^2 + kx^3 + \cdots + kx^n = k(1 + x + x^2 + x^3 + \cdots + x^n).$$

We note that

$$\begin{aligned} s_n(1-x) &= k(1 + x + x^2 + x^3 + \cdots + x^n)(1-x) \\ &= k(1 + x + x^2 + x^3 + \cdots + x^n)1 - k(1 + x + x^2 + x^3 + \cdots + x^{n-1} + x^n)x \\ &= k(1 + x + x^2 + x^3 + \cdots + x^n - x - x^2 - x^3 - \cdots - x^n - x^{n+1}) \\ &= k(1 - x^{n+1}) \end{aligned}$$

so

$$\begin{aligned} s_n(1-x) &= k(1 - x^{n+1}) \\ s_n &= k\frac{1 - x^{n+1}}{1-x}. \end{aligned}$$

If $|x| < 1$, $\lim_{n\to\infty} x^n = 0$ so

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} k\frac{1-x^{n+1}}{1-x} = k\frac{1}{1-x}.$$

Thus, when $|x| < 1$ the geometric series converges to $k/(1-x)$. ♣

When, for example, $k = 1$ and $x = 1/2$:

$$s_n = \frac{1-(1/2)^{n+1}}{1-1/2} = \frac{2^{n+1}-1}{2^n} = 2 - \frac{1}{2^n} \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{1-1/2} = 2.$$

We began the chapter with the series

$$\sum_{n=1}^{\infty} \frac{1}{2^n},$$

namely, the geometric series without the first term 1. Each partial sum of this series is 1 less than the corresponding partial sum for the geometric series, so of course the limit is also one less than the value of the geometric series, that is,

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

It is not hard to see that the following theorem follows from Theorem 9.5.

Theorem 9.20: Series are Linear

Suppose that $\sum a_n$ and $\sum b_n$ are convergent series, and c is a constant. Then

1. $\sum ca_n$ *is convergent and* $\sum ca_n = c\sum a_n$
2. $\sum(a_n + b_n)$ *is convergent and* $\sum(a_n + b_n) = \sum a_n + \sum b_n$.

Note that when c is non-zero, the converse of the first part of this theorem is also true. That is, if $\sum ca_n$ is convergent, then $\sum a_n$ is also convergent; if $\sum ca_n$ converges then $\frac{1}{c}\sum ca_n$ must converge.

On the other hand, the converse of the second part of the theorem is not true. For example, if $a_n = 1$ and $b_n = -1$, then $\sum a_n + \sum b_n = \sum 0 = 0$ converges, but each of $\sum a_n$ and $\sum b_n$ diverges.

In general, the sequence of partial sums s_n is harder to understand and analyze than the sequence of terms a_n, and it is difficult to determine whether series converge and if so to what. The following result will let us deal with some simple cases easily.

Theorem 9.21: Divergence Test

If $\sum a_n$ converges then $\lim_{n\to\infty} a_n = 0$.

Proof. Since $\sum a_n$ converges, $\lim_{n\to\infty} s_n = L$ and $\lim_{n\to\infty} s_{n-1} = L$, because this really says the same thing but "renumbers" the terms. By Theorem 9.5,

$$\lim_{n\to\infty}(s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = L - L = 0.$$

But

$$s_n - s_{n-1} = (a_0 + a_1 + a_2 + \cdots + a_n) - (a_0 + a_1 + a_2 + \cdots + a_{n-1}) = a_n,$$

so as desired $\lim_{n\to\infty} a_n = 0$. ♣

This theorem presents an easy divergence test: Given a series $\sum a_n$, if the limit $\lim_{n\to\infty} a_n$ does not exist or has a value other than zero, the series diverges. Note well that the converse is *not* true: If $\lim_{n\to\infty} a_n = 0$ then the series does not necessarily converge.

Theorem 9.22: The n-th Term Test

If $\lim_{n\to\infty} a_n \neq 0$ or if the limit does not exist, then $\sum a_n$ diverges.

Proof. Consider the statement of the theorem in contrapositive form:

$$\textbf{If } \sum_{n=1}^{\infty} a_n \text{ converges, then } \lim_{n\to\infty} a_n = 0.$$

If s_n are the partial sums of the series, then the assumption that the series converges gives us

$$\lim_{n\to\infty} s_n = s$$

for some number s. Then

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = s - s = 0.$$

♣

Example 9.23:

Show that $\sum_{n=1}^{\infty} \frac{n}{n+1}$ *diverges.*

Solution. We compute the limit:

$$\lim_{n\to\infty} \frac{n}{n+1} = 1 \neq 0.$$

Looking at the first few terms perhaps makes it clear that the series has no chance of converging:

$$\frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \frac{4}{5} + \cdots$$

will just get larger and larger; indeed, after a bit longer the series starts to look very much like $\cdots + 1 + 1 + 1 + 1 + \cdots$, and of course if we add up enough 1's we can make the sum as large as we desire. ♣

Example 9.24: Harmonic Series

Show that $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n}$ *diverges.*

Solution. Here the theorem does not apply: $\lim_{n\to\infty} 1/n = 0$, so it looks like perhaps the series converges. Indeed, if you have the fortitude (or the software) to add up the first 1000 terms you will find that

$$\sum_{n=1}^{1000}\frac{1}{n}\approx 7.49,$$

so it might be reasonable to speculate that the series converges to something in the neighborhood of 10. But in fact the partial sums do go to infinity; they just get big very, very slowly. Consider the following:

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}>1+\frac{1}{2}+\frac{1}{4}+\frac{1}{4}=1+\frac{1}{2}+\frac{1}{2}$$

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}>1+\frac{1}{2}+\frac{1}{4}+\frac{1}{4}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}$$

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{16}>1+\frac{1}{2}+\frac{1}{4}+\frac{1}{4}+\frac{1}{8}+\cdots+\frac{1}{8}+\frac{1}{16}+\cdots+\frac{1}{16}=1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}$$

and so on. By swallowing up more and more terms we can always manage to add at least another $1/2$ to the sum, and by adding enough of these we can make the partial sums as big as we like. In fact, it's not hard to see from this pattern that

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{2^n}>1+\frac{n}{2},$$

so to make sure the sum is over 100, for example, we'd add up terms until we get to around $1/2^{198}$, that is, about $4\cdot 10^{59}$ terms. This series, $\sum(1/n)$, is called the **harmonic series**. ♣

We will often make use of the fact that the first few (e.g. any finite number of) terms in a series are irrelevant when determining whether it will converge. In other words, $\sum_{n=0}^{\infty} a_n$ converges if and only if $\sum_{n=N}^{\infty} a_n$ converges for some $N\ge 1$.

Exercises for 9.2

Exercise 9.2.1 *Explain why* $\displaystyle\sum_{n=1}^{\infty}\frac{n^2}{2n^2+1}$ *diverges.*

Exercise 9.2.2 *Explain why* $\displaystyle\sum_{n=1}^{\infty}\frac{5}{2^{1/n}+14}$ *diverges.*

Exercise 9.2.3 *Explain why* $\displaystyle\sum_{n=1}^{\infty}\frac{3}{n}$ *diverges.*

Exercise 9.2.4 *Compute* $\displaystyle\sum_{n=0}^{\infty}\frac{4}{(-3)^n}-\frac{3}{3^n}$.

Exercise 9.2.5 *Compute* $\displaystyle\sum_{n=0}^{\infty} \frac{3}{2^n} + \frac{4}{5^n}$.

Exercise 9.2.6 *Compute* $\displaystyle\sum_{n=0}^{\infty} \frac{4^{n+1}}{5^n}$.

Exercise 9.2.7 *Compute* $\displaystyle\sum_{n=0}^{\infty} \frac{3^{n+1}}{7^{n+1}}$.

Exercise 9.2.8 *Compute* $\displaystyle\sum_{n=1}^{\infty} \left(\frac{3}{5}\right)^n$.

Exercise 9.2.9 *Compute* $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{5^{n+1}}$.

9.3 The Integral Test

It is generally quite difficult, often impossible, to determine the value of a series exactly. In many cases it is possible at least to determine whether or not the series converges, and so we will spend most of our time on this problem.

If all of the terms a_n in a series are non-negative, then clearly the sequence of partial sums s_n is non-decreasing. This means that if we can show that the sequence of partial sums is bounded, the series must converge. Many useful and interesting series have this property, and they are among the easiest to understand. Let's look at an example.

Example 9.25:

Show that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ *converges.*

Solution. The terms $1/n^2$ are positive and decreasing, and since $\lim_{x\to\infty} 1/x^2 = 0$, the terms $1/n^2$ approach zero. We seek an upper bound for all the partial sums, that is, we want to find a number N so that $s_n \leq N$ for every n. The upper bound is provided courtesy of integration, and is illustrated in figure 9.2.

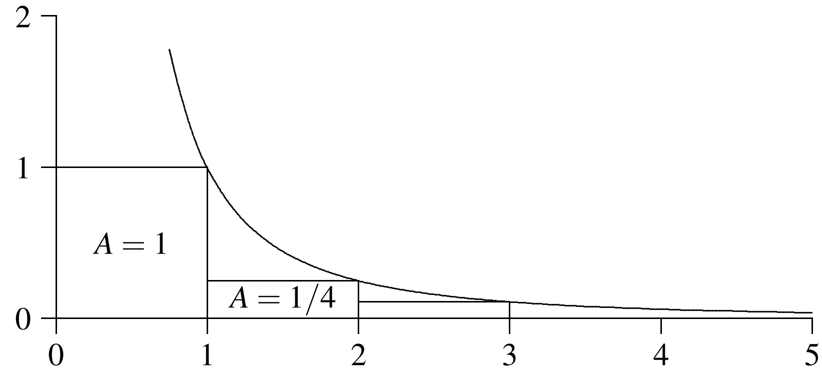


Figure 9.2: Graph of $y = 1/x^2$ with rectangles.

The figure shows the graph of $y = 1/x^2$ together with some rectangles that lie completely below the curve and that all have base length one. Because the heights of the rectangles are determined by the height of the curve, the areas of the rectangles are $1/1^2$, $1/2^2$, $1/3^2$, and so on—in other words, exactly the terms of the series. The partial sum s_n is simply the sum of the areas of the first n rectangles. Because the rectangles all lie between the curve and the x-axis, any sum of rectangle areas is less than the corresponding area under the curve, and so of course any sum of rectangle areas is less than the area under the entire curve. Unfortunately, because of the asymptote at $x = 0$, the integral $\int_0^\infty \frac{1}{x^2}$ is infinite, but we can deal with this by separating the first term from the series and integrating from 1:

$$s_n = \sum_{i=1}^{n} \frac{1}{i^2} = 1 + \sum_{i=2}^{n} \frac{1}{i^2} < 1 + \int_1^n \frac{1}{x^2}\,dx < 1 + \int_1^\infty \frac{1}{x^2}\,dx = 1 + 1 = 2$$

(Recalling that we computed this improper integral in section 7.7). Since the sequence of partial sums s_n is increasing and bounded above by 2, we know that $\lim_{n\to\infty} s_n = L < 2$, and so the series converges to some number less than 2. In fact, it is possible, though difficult, to show that $L = \pi^2/6 \approx 1.6$. ♣

We already know that $\sum 1/n$ diverges. What goes wrong if we try to apply this technique to it? Here's the calculation:

$$s_n = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} < 1 + \int_1^n \frac{1}{x}\,dx < 1 + \int_1^\infty \frac{1}{x}\,dx = 1 + \infty.$$

The problem is that the improper integral doesn't converge. Note that this does *not* prove that $\sum 1/n$ diverges, just that this particular calculation fails to prove that it converges. A slight modification, however, allows us to prove in a second way that $\sum 1/n$ diverges.

Consider a slightly altered version of Figure 9.2, shown in Figure 9.3.

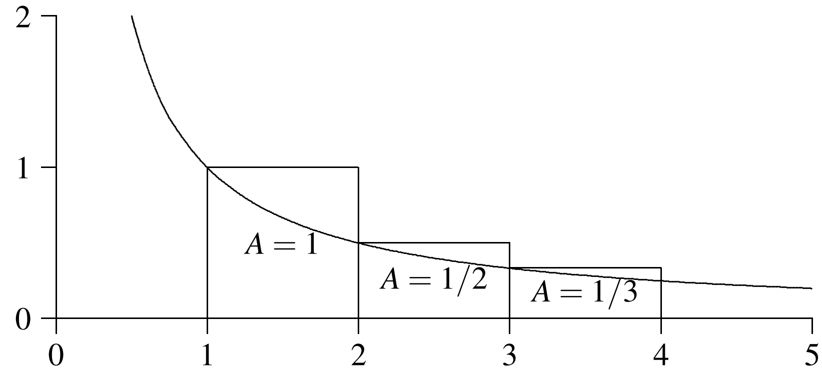


Figure 9.3: Graph of $y = 1/x$ with rectangles.

This time the rectangles are above the curve, that is, each rectangle completely contains the corresponding area under the curve. This means that

$$s_n = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} > \int_1^{n+1} \frac{1}{x}\,dx = \ln x\Big|_1^{n+1} = \ln(n+1).$$

As n gets bigger, $\ln(n+1)$ goes to infinity, so the sequence of partial sums s_n must also go to infinity, so the harmonic series diverges.

The key fact in this example is that

$$\lim_{n\to\infty} \int_1^{n+1} \frac{1}{x}\,dx = \int_1^{\infty} \frac{1}{x}\,dx = \infty$$

So these two examples taken together indicate that we can prove that a series converges or prove that it diverges with a single calculation of an improper integral. This is known as the **integral test**, which we state as a theorem.

Theorem 9.26: Integral Test

Suppose that $f(x) > 0$ and is decreasing on the infinite interval $[1, \infty)$ and that $a_n = f(n)$. Then the series $\sum_{n=1}^{\infty} a_n$ converges if and only if the improper integral $\int_1^{\infty} f(x)\,dx$ converges.

The two examples we have seen are called ***p*-series**; a p-series is any series of the form $\sum 1/n^p$. If $p \le 0$, $\lim_{n\to\infty} 1/n^p \ne 0$, so the series diverges. For positive values of p we can determine precisely which series converge.

Theorem 9.27: p-Series Convergence

A p-series with $p > 0$ converges if and only if $p > 1$.

Proof. We use the integral test; we have already done $p = 1$, so assume that $p \ne 1$.

$$\int_1^{\infty} \frac{1}{x^p}\,dx = \lim_{D\to\infty} \frac{x^{1-p}}{1-p}\Big|_1^D = \lim_{D\to\infty} \frac{D^{1-p}}{1-p} - \frac{1}{1-p}.$$

If $p > 1$ then $1-p < 0$ and $\lim_{D\to\infty} D^{1-p} = 0$, so the integral converges. If $0 < p < 1$ then $1-p > 0$ and $\lim_{D\to\infty} D^{1-p} = \infty$, so the integral diverges. ♣

Example 9.28: *p*-Series

Show that $\sum_{n=1}^{\infty} \frac{1}{n^3}$ *converges.*

Solution. We could of course use the integral test, but now that we have the theorem we may simply note that this is a p-series with $p > 1$. ♣

Example 9.29: *p*-Series

Show that $\sum_{n=1}^{\infty} \frac{5}{n^4}$ *converges.*

Solution. We know that if $\sum_{n=1}^{\infty} 1/n^4$ converges then $\sum_{n=1}^{\infty} 5/n^4$ also converges, by Theorem 9.20. Since $\sum_{n=1}^{\infty} 1/n^4$ is a convergent p-series, $\sum_{n=1}^{\infty} 5/n^4$ converges also. ♣

Example 9.30: *p*-Series

Show that $\sum_{n=1}^{\infty} \frac{5}{\sqrt{n}}$ *diverges.*

Solution. This also follows from Theorem 9.20: Since $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ is a p-series with $p = 1/2 < 1$, it diverges, and so does $\sum_{n=1}^{\infty} \frac{5}{\sqrt{n}}$. ♣

Since it is typically difficult to compute the value of a series exactly, a good approximation is frequently required. In a real sense, a good approximation is only as good as we know it is, that is, while an approximation may in fact be good, it is only valuable in practice if we can guarantee its accuracy to some degree. This guarantee is usually easy to come by for series with decreasing positive terms.

Example 9.31:

Approximate $\sum 1/n^2$ *to within 0.01.*

Solution. Referring to Figure 9.2, if we approximate the sum by $\sum_{n=1}^{N} 1/n^2$, the size of the error we make is the total area of the remaining rectangles, all of which lie under the curve $1/x^2$ from $x = N$ to infinity. So we know the true value of the series is larger than the approximation, and no bigger than the approximation plus the area under the curve from N to infinity. Roughly, then, we need to find N so that

$$\int_N^\infty \frac{1}{x^2}\,dx < 1/100.$$

We can compute the integral:

$$\int_N^\infty \frac{1}{x^2}\,dx = \frac{1}{N},$$

so if we choose $N = 100$ the error will be less than 0.01. Adding up the first 100 terms gives approximately 1.634983900. In fact, we can do a bit better. Since we know that the correct value is between our approximation and our approximation plus the error (not minus), we can cut our error bound in half by taking the value midway between these two values. If we take $N = 50$, we get a sum of 1.6251327 with an error of at most 0.02, so the correct value is between 1.6251327 and 1.6451327, and therefore the value halfway between these, 1.6351327, is within 0.01 of the correct value. We have mentioned that the true value of this series can be shown to be $\pi^2/6 \approx 1.644934068$ which is 0.0098 more than our approximation, and so (just barely) within the required error. Frequently approximations will be even better than the "guaranteed" accuracy, but not always, as this example demonstrates. ♣

Exercises for 9.3

Determine whether each series converges or diverges.

Exercise 9.3.1 $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^{\pi/4}}$

Exercise 9.3.2 $\displaystyle\sum_{n=1}^{\infty} \frac{n}{n^2+1}$

Exercise 9.3.3 $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n^2}$

Exercise 9.3.4 $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2+1}$

Exercise 9.3.5 $\displaystyle\sum_{n=1}^{\infty} \frac{1}{e^n}$

Exercise 9.3.6 $\displaystyle\sum_{n=1}^{\infty} \frac{n}{e^n}$

Exercise 9.3.7 $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n\ln n}$

Exercise 9.3.8 $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^2}$

Exercise 9.3.9 *Find an N so that* $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^4}$ *is between* $\displaystyle\sum_{n=1}^{N} \frac{1}{n^4}$ *and* $\displaystyle\sum_{n=1}^{N} \frac{1}{n^4} + 0.005.$

Exercise 9.3.10 *Find an N so that* $\displaystyle\sum_{n=0}^{\infty} \frac{1}{e^n}$ *is between* $\displaystyle\sum_{n=0}^{N} \frac{1}{e^n}$ *and* $\displaystyle\sum_{n=0}^{N} \frac{1}{e^n} + 10^{-4}.$

Exercise 9.3.11 *Find an N so that* $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n^2}$ *is between* $\displaystyle\sum_{n=1}^{N} \frac{\ln n}{n^2}$ *and* $\displaystyle\sum_{n=1}^{N} \frac{\ln n}{n^2} + 0.005.$

Exercise 9.3.12 *Find an N so that* $\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^2}$ *is between* $\sum_{n=2}^{N} \frac{1}{n(\ln n)^2}$ *and* $\sum_{n=2}^{N} \frac{1}{n(\ln n)^2} + 0.005$.

9.4 Alternating Series

Next we consider series with both positive and negative terms, but in a regular pattern: they alternate, as in the **alternating harmonic series**:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \frac{1}{1} + \frac{-1}{2} + \frac{1}{3} + \frac{-1}{4} + \cdots = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots.$$

In this example the magnitude of the terms decrease, that is, $|a_n|$ forms a decreasing sequence, although this is not required in an alternating series. Recall that for a series with positive terms, if the limit of the terms is not zero, the series cannot converge; but even if the limit of the terms is zero, the series still may not converge. It turns out that for alternating series, the series converges exactly when the limit of the terms is zero. In Figure 9.4, we illustrate what happens to the partial sums of the alternating harmonic series. Because the sizes of the terms a_n are decreasing, the odd partial sums s_1, s_3, s_5, and so on, form a decreasing sequence that is bounded below by s_2, so this sequence must converge. Likewise, the even partial sums s_2, s_4, s_6, and so on, form an increasing sequence that is bounded above by s_1, so this sequence also converges. Since all the even numbered partial sums are less than all the odd numbered ones, and since the "jumps" (that is, the a_i terms) are getting smaller and smaller, the two sequences must converge to the same value, meaning the entire sequence of partial sums $s_1, s_2, s_3, \ldots$ converges as well.

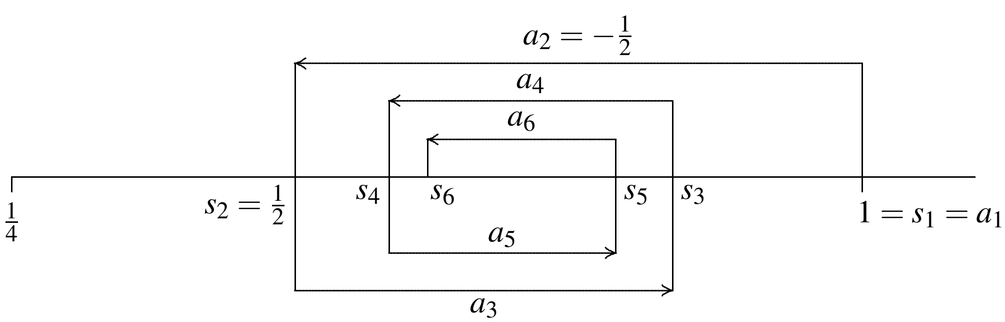


Figure 9.4: The alternating harmonic series.

The same argument works for any alternating sequence with terms that decrease in absolute value. The alternating series test is worth calling a theorem.

Theorem 9.32: Alternating Series Test

Suppose that $\{a_n\}_{n=1}^{\infty}$ *is a non-increasing sequence of positive numbers and* $\lim_{n\to\infty} a_n = 0$. *Then the alternating series* $\sum_{n=1}^{\infty} (-1)^{n-1} a_n$ *converges.*

Proof. The odd-numbered partial sums, $s_1, s_3, s_5, \ldots, s_{2k+1}, \ldots$, form a non-increasing sequence, because $s_{2k+3} = s_{2k+1} - a_{2k+2} + a_{2k+3} \leq s_{2k+1}$, since $a_{2k+2} \geq a_{2k+3}$. This sequence is bounded below by s_2, so it must converge, to some value L. Likewise, the partial sums $s_2, s_4, s_6, \ldots, s_{2k}, \ldots$, form a non-decreasing sequence that is bounded above by s_1, so this sequence also converges, to some value M. Since $\lim_{n\to\infty} a_n = 0$ and $s_{2k+1} = s_{2k} + a_{2k+1}$,

$$L = \lim_{k\to\infty} s_{2k+1} = \lim_{k\to\infty}(s_{2k} + a_{2k+1}) = \lim_{k\to\infty} s_{2k} + \lim_{k\to\infty} a_{2k+1} = M + 0 = M,$$

so $L = M$; the two sequences of partial sums converge to the same limit, and this means the entire sequence of partial sums also converges to L. ♣

Another useful fact is implicit in this discussion. Suppose that

$$L = \sum_{n=1}^{\infty} (-1)^{n-1} a_n$$

and that we approximate L by a finite part of this sum, say

$$L \approx \sum_{n=1}^{N} (-1)^{n-1} a_n.$$

Because the terms are decreasing in size, we know that the true value of L must be between this approximation and the next one, that is, between

$$\sum_{n=1}^{N} (-1)^{n-1} a_n \quad \text{and} \quad \sum_{n=1}^{N+1} (-1)^{n-1} a_n.$$

Depending on whether N is odd or even, the second will be larger or smaller than the first.

Example 9.33:

Approximate the sum of the alternating harmonic series to within 0.05.

Solution. We need to go to the point at which the next term to be added or subtracted is $1/10$. Adding up the first nine and the first ten terms we get approximately 0.746 and 0.646. These are $1/10$ apart, so the value halfway between them, 0.696, is within 0.05 of the correct value. ♣

We have considered alternating series with first index 1, and in which the first term is positive, but a little thought shows this is not crucial. The same test applies to any similar series, such as $\sum_{n=0}^{\infty} (-1)^n a_n$, $\sum_{n=1}^{\infty} (-1)^n a_n$, $\sum_{n=17}^{\infty} (-1)^n a_n$, etc.

Exercises for 9.4

Determine whether the following series converge or diverge.

Exercise 9.4.1 $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n+5}$

Exercise 9.4.2 $\displaystyle\sum_{n=4}^{\infty} \frac{(-1)^{n-1}}{\sqrt{n-3}}$

Exercise 9.4.3 $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{n}{3n-2}$

Exercise 9.4.4 $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{\ln n}{n}$

Exercise 9.4.5 *Approximate* $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{1}{n^3}$ *to within 0.005.*

Exercise 9.4.6 *Approximate* $\displaystyle\sum_{n=1}^{\infty} (-1)^{n-1}\frac{1}{n^4}$ *to within 0.005.*

9.5 Comparison Tests

As we begin to compile a list of convergent and divergent series, new ones can sometimes be analyzed by comparing them to ones that we already understand.

Example 9.34:

Does $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n^2 \ln n}$ *converge?*

Solution. The obvious first approach, based on what we know, is the integral test. Unfortunately, we can't compute the required antiderivative. But looking at the series, it would appear that it must converge, because the terms we are adding are smaller than the terms of a p-series, that is,

$$\frac{1}{n^2 \ln n} < \frac{1}{n^2},$$

when $n \geq 3$. Since adding up the terms $1/n^2$ doesn't get "too big", the new series "should" also converge. Let's make this more precise.

The series $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n^2 \ln n}$ converges if and only if $\displaystyle\sum_{n=3}^{\infty} \frac{1}{n^2 \ln n}$ converges—all we've done is dropped the initial term. We know that $\displaystyle\sum_{n=3}^{\infty} \frac{1}{n^2}$ converges. Looking at two typical partial sums:

$$s_n = \frac{1}{3^2 \ln 3} + \frac{1}{4^2 \ln 4} + \frac{1}{5^2 \ln 5} + \cdots + \frac{1}{n^2 \ln n} < \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots + \frac{1}{n^2} = t_n.$$

Since the p-series converges, say to L, and since the terms are positive, $t_n < L$. Since the terms of the new series are positive, the s_n form an increasing sequence and $s_n < t_n < L$ for all n. Hence the sequence $\{s_n\}$ is bounded and so converges. ♣

Sometimes, even when the integral test applies, comparison to a known series is easier, so it's generally a good idea to think about doing a comparison before doing the integral test.

Example 9.35:

Does $\sum_{n=2}^{\infty} \frac{|\sin n|}{n^2}$ *converge?*

Solution. We can't apply the integral test here, because the terms of this series are not decreasing. Just as in the previous example, however,

$$\frac{|\sin n|}{n^2} \leq \frac{1}{n^2},$$

because $|\sin n| \leq 1$. Once again the partial sums are non-decreasing and bounded above by $\sum 1/n^2 = L$, so the new series converges. ♣

Like the integral test, the comparison test can be used to show both convergence and divergence. In the case of the integral test, a single calculation will confirm whichever is the case. To use the comparison test we must first have a good idea as to convergence or divergence and pick the sequence for comparison accordingly.

Example 9.36:

Does $\sum_{n=2}^{\infty} \frac{1}{\sqrt{n^2-3}}$ *converge?*

Solution. We observe that the -3 should have little effect compared to the n^2 inside the square root, and therefore guess that the terms are enough like $1/\sqrt{n^2} = 1/n$ that the series should diverge. We attempt to show this by comparison to the harmonic series. We note that

$$\frac{1}{\sqrt{n^2-3}} > \frac{1}{\sqrt{n^2}} = \frac{1}{n},$$

so that

$$s_n = \frac{1}{\sqrt{2^2-3}} + \frac{1}{\sqrt{3^2-3}} + \cdots + \frac{1}{\sqrt{n^2-3}} > \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = t_n,$$

where t_n is 1 less than the corresponding partial sum of the harmonic series (because we start at $n = 2$ instead of $n = 1$). Since $\lim_{n\to\infty} t_n = \infty$, $\lim_{n\to\infty} s_n = \infty$ as well. ♣

So the general approach is this: If you believe that a new series is convergent, attempt to find a convergent series whose terms are larger than the terms of the new series; if you believe that a new series is divergent, attempt to find a divergent series whose terms are smaller than the terms of the new series.

Example 9.37:

Does $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^2+3}}$ *converge?*

Solution. Just as in the last example, we guess that this is very much like the harmonic series and so diverges. Unfortunately,

$$\frac{1}{\sqrt{n^2+3}} < \frac{1}{n},$$

so we can't compare the series directly to the harmonic series. A little thought leads us to

$$\frac{1}{\sqrt{n^2+3}} > \frac{1}{\sqrt{n^2+3n^2}} = \frac{1}{2n},$$

so if $\sum 1/(2n)$ diverges then the given series diverges. But since $\sum 1/(2n) = (1/2)\sum 1/n$, Theorem 9.20 implies that it does indeed diverge. ♣

For reference we summarize the comparison test in a theorem.

Theorem 9.38: Comparison Theorem

Suppose that a_n and b_n are non-negative for all n and that $a_n \le b_n$ when $n \ge N$, for some N.

- *If $\sum_{n=0}^{\infty} b_n$ converges, so does $\sum_{n=0}^{\infty} a_n$.*
- *If $\sum_{n=0}^{\infty} a_n$ diverges, so does $\sum_{n=0}^{\infty} b_n$.*

Exercises for 9.5

Determine whether the series converge or diverge.

Exercise 9.5.1 $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n^2+3n+5}$

Exercise 9.5.2 $\displaystyle\sum_{n=2}^{\infty} \frac{1}{2n^2+3n-5}$

Exercise 9.5.3 $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n^2-3n-5}$

Exercise 9.5.4 $\displaystyle\sum_{n=1}^{\infty} \frac{3n+4}{2n^2+3n+5}$

Exercise 9.5.5 $\displaystyle\sum_{n=1}^{\infty} \frac{3n^2+4}{2n^2+3n+5}$

Exercise 9.5.6 $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n}$

Exercise 9.5.7 $\displaystyle\sum_{n=1}^{\infty} \frac{\ln n}{n^3}$

Exercise 9.5.8 $\displaystyle\sum_{n=2}^{\infty} \frac{1}{\ln n}$

Exercise 9.5.9 $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{2^n+5^n}$

Exercise 9.5.10 $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{2^n+3^n}$

9.6 Absolute Convergence

Roughly speaking there are two ways for a series to converge: As in the case of $\sum 1/n^2$, the individual terms get small very quickly, so that the sum of all of them stays finite, or, as in the case of $\sum(-1)^{n-1}/n$, the terms don't get small fast enough ($\sum 1/n$ diverges), but a mixture of positive and negative terms provides enough cancellation to keep the sum finite. You might guess from what we've seen that if the terms get small fast enough that the sum of their absolute values converges, then the series will still converge regardless of which terms are actually positive or negative.

Theorem 9.39: Absolute Convergence

If $\sum_{n=0}^{\infty} |a_n|$ *converges, then* $\sum_{n=0}^{\infty} a_n$ *converges.*

Proof. Note that $0 \le a_n + |a_n| \le 2|a_n|$ so by the comparison test $\sum_{n=0}^{\infty}(a_n + |a_n|)$ converges. Now

$$\sum_{n=0}^{\infty}(a_n + |a_n|) - \sum_{n=0}^{\infty}|a_n| = \sum_{n=0}^{\infty} a_n + |a_n| - |a_n| = \sum_{n=0}^{\infty} a_n$$

converges by Theorem 9.20. ♣

So given a series $\sum a_n$ with both positive and negative terms, you should first ask whether $\sum |a_n|$ converges. This may be an easier question to answer, because we have tests that apply specifically to terms with non-negative terms. If $\sum |a_n|$ converges then you know that $\sum a_n$ converges as well. If $\sum |a_n|$ diverges then it still may be true that $\sum a_n$ converges, but you will need to use other techniques to decide. Intuitively this results says that it is (potentially) easier for $\sum a_n$ to converge than for $\sum |a_n|$ to converge, because terms may partially cancel in the first series.

If $\sum |a_n|$ converges we say that $\sum a_n$ converges **absolutely**; to say that $\sum a_n$ converges absolutely is to say that the terms of the series get small (in absolute value) quickly enough to guarantee that the series converges, regardless of whether any of the terms cancel each other. If $\sum a_n$ converges but $\sum |a_n|$ does not, we say that $\sum a_n$ converges **conditionally**. For example $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n^2}$ converges absolutely, while $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}$ converges conditionally.

Example 9.40:

Does $\sum_{n=2}^{\infty}\frac{\sin n}{n^2}$ *converge?*

Solution. In Example 9.35 we saw that $\sum_{n=2}^{\infty}\frac{|\sin n|}{n^2}$ converges, so the given series converges absolutely. ♣

Example 9.41:

Does $\sum_{n=0}^{\infty}(-1)^n\frac{3n+4}{2n^2+3n+5}$ *converge?*

Solution. Taking the absolute value, $\sum_{n=0}^{\infty}\frac{3n+4}{2n^2+3n+5}$ diverges by comparison to $\sum_{n=1}^{\infty}\frac{3}{10n}$, so if the series converges it does so conditionally. It is true that $\lim_{n\to\infty}(3n+4)/(2n^2+3n+5)=0$, so to apply the alternating series test we need to know whether the terms are decreasing. If we let $f(x)=(3x+4)/(2x^2+3x+5)$ then $f'(x)=-(6x^2+16x-3)/(2x^2+3x+5)^2$, and it is not hard to see that this is negative for $x\geq 1$, so the series is decreasing and by the alternating series test it converges. ♣

Exercises for 9.6

Determine whether each series converges absolutely, converges conditionally, or diverges.

Exercise 9.6.1 $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{2n^2+3n+5}$

Exercise 9.6.2 $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{3n^2+4}{2n^2+3n+5}$

Exercise 9.6.3 $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\ln n}{n}$

Exercise 9.6.4 $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\ln n}{n^3}$

Exercise 9.6.5 $\sum_{n=2}^{\infty}(-1)^{n}\frac{1}{\ln n}$

Exercise 9.6.6 $\sum_{n=0}^{\infty}(-1)^{n}\frac{3^n}{2^n+5^n}$

Exercise 9.6.7 $\sum_{n=0}^{\infty}(-1)^{n}\frac{3^n}{2^n+3^n}$

Exercise 9.6.8 $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\arctan n}{n}$

9.7 The Ratio and Root Tests

Does the series $\sum_{n=0}^{\infty}\frac{n^5}{5^n}$ converge? It is possible, but a bit unpleasant, to approach this with the integral test or the comparison test, but there is an easier way. Consider what happens as we move from one term to the next in this series:

$$\cdots+\frac{n^5}{5^n}+\frac{(n+1)^5}{5^{n+1}}+\cdots$$

The denominator goes up by a factor of 5, $5^{n+1} = 5 \cdot 5^n$, but the numerator goes up by much less: $(n+1)^5 = n^5 + 5n^4 + 10n^3 + 10n^2 + 5n + 1$, which is much less than $5n^5$ when n is large, because $5n^4$ is much less than n^5. So we might guess that in the long run it begins to look as if each term is $1/5$ of the previous term. We have seen series that behave like this: The geometric series.

$$\sum_{n=0}^{\infty} \frac{1}{5^n} = \frac{5}{4},$$

So we might try comparing the given series to some variation of this geometric series. This is possible, but a bit messy. We can in effect do the same thing, but bypass most of the unpleasant work.

The key is to notice that

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{(n+1)^5}{5^{n+1}} \frac{5^n}{n^5} = \lim_{n\to\infty} \frac{(n+1)^5}{n^5} \frac{1}{5} = 1 \cdot \frac{1}{5} = \frac{1}{5}.$$

This is really just what we noticed above, done a bit more formally: in the long run, each term is one fifth of the previous term. Now pick some number between $1/5$ and 1, say $1/2$. Because

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \frac{1}{5},$$

then when n is big enough, say $n \geq N$ for some N,

$$\frac{a_{n+1}}{a_n} < \frac{1}{2} \quad \text{so} \quad a_{n+1} < \frac{a_n}{2}.$$

So $a_{N+1} < a_N/2$, $a_{N+2} < a_{N+1}/2 < a_N/4$, $a_{N+3} < a_{N+2}/2 < a_N/8$, and so on. The general form is $a_{N+k} < a_N/2^k$. So if we look at the series

$$\sum_{k=0}^{\infty} a_{N+k} = a_N + a_{N+1} + a_{N+2} + a_{N+3} + \cdots + a_{N+k} + \cdots,$$

its terms are less than or equal to the terms of the sequence

$$a_N + \frac{a_N}{2} + \frac{a_N}{4} + \frac{a_N}{8} + \cdots + \frac{a_N}{2^k} + \cdots = \sum_{k=0}^{\infty} \frac{a_N}{2^k} = 2a_N.$$

So by the comparison test, $\sum_{k=0}^{\infty} a_{N+k}$ converges, and this means that $\sum_{n=0}^{\infty} a_n$ converges, since we've just added the fixed number $a_0 + a_1 + \cdots + a_{N-1}$.

Under what circumstances could we do this? What was crucial was that the limit of a_{n+1}/a_n, say L, was less than 1 so that we could pick a value r so that $L < r < 1$. The fact that $L < r$ ($1/5 < 1/2$ in our example) means that we can compare the series $\sum a_n$ to $\sum r^n$, and the fact that $r < 1$ guarantees that $\sum r^n$ converges. That's really all that is required to make the argument work. We also made use of the fact that the terms of the series were positive; in general we simply consider the absolute values of the terms and we end up testing for absolute convergence.

Theorem 9.42: The Ratio Test

Suppose that $\lim_{n\to\infty} |a_{n+1}/a_n| = L$. *If* $L < 1$ *the series* $\sum a_n$ *converges absolutely, if* $L > 1$ *the series diverges, and if* $L = 1$ *this test gives no information.*

Proof. The example above essentially proves the first part of this, if we simply replace $1/5$ by L and $1/2$ by r. Suppose that $L > 1$, and pick r so that $1 < r < L$. Then for $n \geq N$, for some N,

$$\frac{|a_{n+1}|}{|a_n|} > r \quad \text{and} \quad |a_{n+1}| > r|a_n|.$$

This implies that $|a_{N+k}| > r^k|a_N|$, but since $r > 1$ this means that $\lim_{k\to\infty} |a_{N+k}| \neq 0$, which means also that $\lim_{n\to\infty} a_n \neq 0$. By the divergence test, the series diverges.

To see that we get no information when $L = 1$, we need to exhibit two series with $L = 1$, one that converges and one that diverges. The series $\sum 1/n^2$ and $\sum 1/n$ provide a simple example. ♣

The ratio test is particularly useful for series involving the factorial function.

Example 9.43:

Analyze $\displaystyle\sum_{n=0}^{\infty} \frac{5^n}{n!}$.

Solution.

$$\lim_{n\to\infty} \frac{5^{n+1}}{(n+1)!}\frac{n!}{5^n} = \lim_{n\to\infty} \frac{5^{n+1}}{5^n}\frac{n!}{(n+1)!} = \lim_{n\to\infty} 5\frac{1}{(n+1)} = 0.$$

Since $0 < 1$, the series converges. ♣

A similar argument justifies a similar test that is occasionally easier to apply.

Theorem 9.44: The Root Test

Suppose that $\lim_{n\to\infty} |a_n|^{1/n} = L$. *If* $L < 1$ *the series* $\sum a_n$ *converges absolutely, if* $L > 1$ *the series diverges, and if* $L = 1$ *this test gives no information.*

The proof of the root test is actually easier than that of the ratio test, and is left as an exercise.

Example 9.45:

Analyze $\displaystyle\sum_{n=0}^{\infty} \frac{5^n}{n^n}$.

Solution. The ratio test turns out to be a bit difficult on this series (try it). Using the root test:

$$\lim_{n\to\infty} \left(\frac{5^n}{n^n}\right)^{1/n} = \lim_{n\to\infty} \frac{(5^n)^{1/n}}{(n^n)^{1/n}} = \lim_{n\to\infty} \frac{5}{n} = 0.$$

Since $0 < 1$, the series converges. ♣

The root test is frequently useful when n appears as an exponent in the general term of the series.

Exercises for 9.7

Exercise 9.7.1 *Compute* $\lim_{n\to\infty} |a_{n+1}/a_n|$ *for the series* $\sum 1/n^2$.

Exercise 9.7.2 *Compute* $\lim_{n\to\infty} |a_{n+1}/a_n|$ *for the series* $\sum 1/n$.

Exercise 9.7.3 *Compute* $\lim_{n\to\infty} |a_n|^{1/n}$ *for the series* $\sum 1/n^2$.

Exercise 9.7.4 *Compute* $\lim_{n\to\infty} |a_n|^{1/n}$ *for the series* $\sum 1/n$.

Exercise 9.7.5 *Determine whether the series converge.*

(a) $\displaystyle\sum_{n=0}^{\infty} (-1)^n \frac{3^n}{5^n}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{n^n}$

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{n^5}{n^n}$

(d) $\displaystyle\sum_{n=1}^{\infty} \frac{(n!)^2}{n^n}$

Exercise 9.7.6 *Prove Theorem 9.44, the root test.*

9.8 Power Series

Recall that the sum of a geometric series can be expressed using the simple formula:

$$\sum_{n=0}^{\infty} kx^n = \frac{k}{1-x},$$

if $|x| < 1$, and that the series diverges when $|x| \geq 1$. At the time, we thought of x as an unspecified constant, but we could just as well think of it as a variable, in which case the series

$$\sum_{n=0}^{\infty} kx^n$$

is a function, namely, the function $k/(1-x)$, as long as $|x| < 1$: Looking at this from the opposite perspective, this means that the function $k/(1-x)$ can be represented as the sum of an infinite series. Why would this be useful? While $k/(1-x)$ is a reasonably easy function to deal with, the more complicated representation $\sum kx^n$ does have some advantages: it appears to be an infinite version of one of the simplest function types—a polynomial. Later on we will investigate some of the ways we can take advantage of this 'infinite polynomial' representation, but first we should ask if other functions can even be represented this way.

The geometric series has a special feature that makes it unlike a typical polynomial—the coefficients of the powers of x are all the same, namely k. We will need to allow more general coefficients if we are to get anything other than the geometric series.

Definition 9.46: Power Series

A power series is a series of the form

$$\sum_{n=0}^{\infty} a_n x^n,$$

where each a_n is a real number.

As we did in the section on sequences, we can think of the a_n as being a function $a(n)$ defined on the non-negative integers. Note, however, that the a_n do not depend on x.

Example 9.47:

Determine whether the power series $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n}$ converges.

Solution. We can investigate convergence using the ratio test:

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

Thus when $|x| < 1$ the series converges and when $|x| > 1$ it diverges, leaving only two values in doubt. When $x = 1$ the series is the harmonic series and diverges; when $x = -1$ it is the alternating harmonic series (actually the negative of the usual alternating harmonic series) and converges. Thus, we may think of $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n}$ as a function from the interval $[-1,1)$ to the real numbers. ♣

A bit of thought reveals that the ratio test applied to a power series will always have the same nice form. In general, we will compute

$$\lim_{n\to\infty} \frac{|a_{n+1}||x|^{n+1}}{|a_n||x|^n} = \lim_{n\to\infty} |x| \frac{|a_{n+1}|}{|a_n|} = |x| \lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L|x|,$$

assuming that $\lim |a_{n+1}|/|a_n|$ exists. Then the series converges if $L|x| < 1$, that is, if $|x| < 1/L$, and diverges if $|x| > 1/L$. Only the two values $x = \pm 1/L$ require further investigation. Thus the series will always define a function on the interval $(-1/L, 1/L)$, that perhaps will extend to one or both endpoints as well. Two special cases deserve mention: if $L = 0$ the limit is 0 no matter what value x takes, so the series converges for all x and the function is defined for all real numbers. If $L = \infty$, then no matter what value x

takes the limit is infinite and the series converges only when $x = 0$. The value $1/L$ is called the **radius of convergence** of the series, and the interval on which the series converges is the **interval of convergence**.

We can make these ideas a bit more general. Consider the series

$$\sum_{n=0}^{\infty} \frac{(x+2)^n}{3^n}$$

This looks a lot like a power series, but with $(x+2)^n$ instead of x^n. Let's try to determine the values of x for which it converges. This is just a geometric series, so it converges when

$$\begin{aligned} |x+2|/3 &< 1 \\ |x+2| &< 3 \\ -3 < x+2 &< 3 \\ -5 < x &< 1. \end{aligned}$$

So the interval of convergence for this series is $(-5, 1)$. The center of this interval is at -2, which is at distance 3 from the endpoints, so the radius of convergence is 3, and we say that the series is centered at -2.

Interestingly, if we compute the sum of the series we get

$$\sum_{n=0}^{\infty} \left(\frac{x+2}{3}\right)^n = \frac{1}{1-\frac{x+2}{3}} = \frac{3}{1-x}.$$

Multiplying both sides by 1/3 we obtain

$$\sum_{n=0}^{\infty} \frac{(x+2)^n}{3^{n+1}} = \frac{1}{1-x},$$

which we recognize as being equal to

$$\sum_{n=0}^{\infty} x^n,$$

so we have two series with the same sum but different intervals of convergence.

This leads to the following definition:

Definition 9.48: Power Series

A power series centered at c has the form

$$\sum_{n=0}^{\infty} a_n (x-c)^n,$$

where c and each a_n are real numbers.

Exercises for 9.8

Exercise 9.8.1 *Find the radius and interval of convergence for each series. In part c), do not attempt to determine whether the endpoints are in the interval of convergence.*

(a) $\displaystyle\sum_{n=0}^{\infty} nx^n$

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

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{n^n}(x-2)^n$

(d) $\displaystyle\sum_{n=1}^{\infty} \frac{(n!)^2}{n^n}(x-2)^n$

(e) $\displaystyle\sum_{n=1}^{\infty} \frac{(x+5)^n}{n(n+1)}$

Exercise 9.8.2 *Find the radius of convergence for the series* $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{n^n}x^n$.

9.9 Calculus with Power Series

We now know that some functions can be expressed as power series, which look like infinite polynomials. Since it is easy to find derivatives and integrals of polynomials, we might hope that we can take derivatives and integrals of power series in an analogous way. In fact we can, as stated in the following theorem, which we will not prove here.

Theorem 9.49:

Suppose the power series $f(x) = \displaystyle\sum_{n=0}^{\infty} a_n(x-a)^n$ *has radius of convergence* R. *Then*

$$f'(x) = \sum_{n=0}^{\infty} na_n(x-a)^{n-1},$$

$$\int f(x)\,dx = C + \sum_{n=0}^{\infty} \frac{a_n}{n+1}(x-a)^{n+1},$$

and these two series have radius of convergence R.

Example 9.50:

Find a power series representation of $\ln|1-x|$.

Solution. Starting with the geometric series:

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

$$\int \frac{1}{1-x}\,dx = -\ln|1-x| = \sum_{n=0}^{\infty} \frac{1}{n+1}x^{n+1}$$

$$\ln|1-x| = \sum_{n=0}^{\infty} -\frac{1}{n+1}x^{n+1}$$

when $|x| < 1$. The series does not converge when $x = 1$ but does converge when $x = -1$ or $1-x = 2$. The interval of convergence is $[-1,1)$, or $0 < 1-x \le 2$. We can use this series to express $\ln(a)$ as a series when $0 < a \le 2$ by setting $x-1 = a$. For example

$$\ln(3/2) = \ln(1-(-1/2)) = \sum_{n=0}^{\infty} (-1)^n \frac{1}{n+1}\frac{1}{2^{n+1}}.$$

We can use this in turn to approximate $\ln(3/2)$:

$$\ln(3/2) \approx \frac{1}{2} - \frac{1}{8} + \frac{1}{24} - \frac{1}{64} + \frac{1}{160} - \frac{1}{384} + \frac{1}{896} = \frac{909}{2240} \approx 0.406.$$

Because this is an alternating series with decreasing terms, we know that the true value is between $909/2240$ and $909/2240 - 1/2048 = 29053/71680 \approx .4053$, so $0.4053 \le \ln(3/2) \le 0.406$. ♣

With a bit of arithmetic, we can approximate values outside of the interval of convergence:

Example 9.51:

Find an approximation for $\ln(9/4)$.

Solution. We can use the approximation we just computed, plus some rules for logarithms:

$$\ln(9/4) = \ln((3/2)^2) = 2\ln(3/2) \approx 0.812,$$

and using our bounds above,

$$0.8106 \le \ln(9/4) \le 0.812.$$

♣

Exercises for 9.9

Exercise 9.9.1 *Find a series representation for* $\ln 2$.

Exercise 9.9.2 *Find a power series representation for* $1/(1-x)^2$.

Exercise 9.9.3 *Find a power series representation for* $2/(1-x)^3$.

Exercise 9.9.4 *Find a power series representation for* $1/(1-x)^3$. *What is the radius of convergence?*

Exercise 9.9.5 *Find a power series representation for* $\int \ln(1-x)\,dx$.

9.10 Taylor Series

We have seen that some functions can be represented as series, which may give valuable information about the function. So far, we have seen only those examples that result from manipulation of our one fundamental example, the geometric series. We would like to start with a given function and produce a series to represent it, if possible.

Suppose that $f(x) = \sum_{n=0}^{\infty} a_n x^n$ on some interval of convergence centered at 0. Then we know that we can compute derivatives of f by taking derivatives of the terms of the series. Let's look at the first few in general:

$$\begin{aligned}
f'(x) &= \sum_{n=1}^{\infty} n a_n x^{n-1} = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots \\
f''(x) &= \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} = 2a_2 + 3\cdot 2a_3 x + 4\cdot 3a_4 x^2 + \cdots \\
f'''(x) &= \sum_{n=3}^{\infty} n(n-1)(n-2) a_n x^{n-3} = 3\cdot 2a_3 + 4\cdot 3\cdot 2a_4 x + \cdots
\end{aligned}$$

By examining these it's not hard to discern the general pattern. The kth derivative must be

$$\begin{aligned}
f^{(k)}(x) &= \sum_{n=k}^{\infty} n(n-1)(n-2)\cdots(n-k+1) a_n x^{n-k} \\
&= k(k-1)(k-2)\cdots(2)(1)a_k + (k+1)(k)\cdots(2)a_{k+1}x + \\
&\quad + (k+2)(k+1)\cdots(3)a_{k+2}x^2 + \cdots
\end{aligned}$$

We can express this more clearly by using factorial notation:

$$f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n x^{n-k} = k! a_k + (k+1)! a_{k+1} x + \frac{(k+2)!}{2!} a_{k+2} x^2 + \cdots$$

We can solve for a_n by substituting $x = 0$ in the formula for $f^{(k)}(x)$:

$$f^{(k)}(0) = k! a_k + \sum_{n=k+1}^{\infty} \frac{n!}{(n-k)!} a_n 0^{n-k} = k! a_k,$$

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

Note that the original series for f yields $f(0) = a_0$.

So if a function f can be represented by a series, we can easily find such a series. Given a function f, the series

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

is called the **Maclaurin series** for f.

Example 9.52: Maclaurin Series

Find the Maclaurin series for $f(x) = 1/(1-x)$.

Solution. We need to compute the derivatives of f (and hope to spot a pattern).

$$\begin{aligned} f(x) &= (1-x)^{-1} \\ f'(x) &= (1-x)^{-2} \\ f''(x) &= 2(1-x)^{-3} \\ f'''(x) &= 6(1-x)^{-4} \\ f^{(4)}(x) &= 4!(1-x)^{-5} \\ &\vdots \\ f^{(n)}(x) &= n!(1-x)^{-n-1} \end{aligned}$$

So

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{n!(1-0)^{-n-1}}{n!} = 1$$

and the Maclaurin series is

$$\sum_{n=0}^{\infty} 1 \cdot x^n = \sum_{n=0}^{\infty} x^n,$$

the geometric series. ♣

A warning is in order here. Given a function f we may be able to compute the Maclaurin series, but that does not mean we have found a series representation for f. We still need to know where the series converges, and if, where it converges, it converges to $f(x)$. While for most commonly encountered functions the Maclaurin series does indeed converge to f on some interval, this is not true of all functions, so care is required.

As a practical matter, if we are interested in using a series to approximate a function, we will need some finite number of terms of the series. Even for functions with messy derivatives we can compute these using computer software like Sage. If we want to describe a series completely, we would like to be able to write down a formula for a typical term in the series. Fortunately, a few of the most important functions are very easy.

Example 9.53: Maclaurin Series

Find the Maclaurin series for $\sin x$.

Solution. Computing the first few derivatives is simple: $f'(x) = \cos x$, $f''(x) = -\sin x$, $f'''(x) = -\cos x$, $f^{(4)}(x) = \sin x$, and then the pattern repeats. The values of the derivative when $x = 0$ are: 1, 0, -1, 0, 1, 0, -1, 0,..., and so the Maclaurin series is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

We should always determine the radius of convergence:

$$\lim_{n\to\infty} \frac{|x|^{2n+3}}{(2n+3)!} \frac{(2n+1)!}{|x|^{2n+1}} = \lim_{n\to\infty} \frac{|x|^2}{(2n+3)(2n+2)} = 0,$$

so the series converges for every x. Since it turns out that this series does indeed converge to $\sin x$ everywhere, we have a series representation for $\sin x$ for every x. ♣

Sometimes the formula for the nth derivative of a function f is difficult to discover, but a combination of a known Maclaurin series and some algebraic manipulation leads easily to the Maclaurin series for f.

Example 9.54: Maclaurin Series

Find the Maclaurin series for $x\sin(-x)$.

Solution. To get from $\sin x$ to $x\sin(-x)$ we substitute $-x$ for x and then multiply by x. We can do the same thing to the series for $\sin x$:

$$x\sum_{n=0}^{\infty} (-1)^n \frac{(-x)^{2n+1}}{(2n+1)!} = x\sum_{n=0}^{\infty} (-1)^n (-1)^{2n+1} \frac{x^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{x^{2n+2}}{(2n+1)!}.$$

♣

As we have seen, a power series can be centered at a point other than zero, and the method that produces the Maclaurin series can also produce such series.

Example 9.55: Taylor Series

Find a series centered at -2 *for* $1/(1-x)$.

Solution. If the series is $\sum_{n=0}^{\infty} a_n(x+2)^n$ then looking at the kth derivative:

$$k!(1-x)^{-k-1} = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n (x+2)^{n-k}$$

and substituting $x = -2$ we get $k!3^{-k-1} = k!a_k$ and $a_k = 3^{-k-1} = 1/3^{k+1}$, so the series is

$$\sum_{n=0}^{\infty} \frac{(x+2)^n}{3^{n+1}}.$$

♣

Such a series is called the **Taylor series** for the function, and the general term has the form

$$\frac{f^{(n)}(a)}{n!}(x-a)^n.$$

A Maclaurin series is simply a Taylor series with $a = 0$.

Exercises for 9.10

Exercise 9.10.1 *For each function, find the Maclaurin series or Taylor series centered at a, and the radius of convergence.*

(a) $\cos x$

(b) e^x

(c) $1/x$, $a = 5$

(d) $\ln x$, $a = 1$

(e) $\ln x$, $a = 2$

(f) $1/x^2$, $a = 1$

(g) $1/\sqrt{1-x}$

(h) *Find the first four terms of the Maclaurin series for* $\tan x$ *(up to and including the* x^3 *term).*

(i) *Use a combination of Maclaurin series and algebraic manipulation to find a series centered at zero for* $x\cos(x^2)$.

(j) *Use a combination of Maclaurin series and algebraic manipulation to find a series centered at zero for* xe^{-x}.

9.11 Taylor's Theorem

One of the most important uses of infinite series is using an initial portion of the series for f to approximate f. We have seen, for example, that when we add up the first n terms of an alternating series with decreasing terms that the difference between this and the true value is at most the size of the next term. A similar result is true of many Taylor series.

Theorem 9.56:

Suppose that f is defined on some open interval I around a and suppose $f^{(N+1)}(x)$ exists on this interval. Then for each $x \neq a$ in I there is a value z between x and a so that

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1}.$$

Proof. The proof requires some cleverness to set up, but then the details are quite elementary. We define a function $F(t)$ as follows:

$$F(t) = \sum_{n=0}^{N} \frac{f^{(n)}(t)}{n!}(x-t)^n + B(x-t)^{N+1}.$$

Here we have replaced a by t in the first $N+1$ terms of the Taylor series, and added a carefully chosen term on the end, with B to be determined. Note that we are temporarily keeping x fixed, so the only variable in this equation is t, and we will be interested only in t between a and x. Now substitute $t = a$:

$$F(a) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + B(x-a)^{N+1}.$$

Set this equal to $f(x)$:

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + B(x-a)^{N+1}.$$

Since $x \neq a$, we can solve this for B, which is a "constant"—it depends on x and a but those are temporarily fixed. Now we have defined a function $F(t)$ with the property that $F(a) = f(x)$. Also, all terms with a positive power of $(x-t)$ become zero when we substitute x for t, so $F(x) = f^{(0)}(x)/0! = f(x)$. So $F(a) = F(x)$. By Rolle's theorem (5.21), we know that there is a value $z \in (a,x)$ such that $F'(z) = 0$. But what is F'? Each term in $F(t)$, except the first term and the extra term involving B, is a product, so to take the derivative we use the product rule on each of these terms.

$$\begin{aligned} F(t) = f(t) &+ \frac{f^{(1)}(t)}{1!}(x-t)^1 + \frac{f^{(2)}(t)}{2!}(x-t)^2 + \frac{f^{(3)}(t)}{3!}(x-t)^3 + \cdots \\ &+ \frac{f^{(N)}(t)}{N!}(x-t)^N + B(x-t)^{N+1}. \end{aligned}$$

So the derivative is

$$F'(t) = f'(t) + \left(\frac{f^{(1)}(t)}{1!}(x-t)^0(-1) + \frac{f^{(2)}(t)}{1!}(x-t)^1 \right)$$

$$
\begin{aligned}
&+\left(\frac{f^{(2)}(t)}{1!}(x-t)^1(-1)+\frac{f^{(3)}(t)}{2!}(x-t)^2\right)\\
&+\left(\frac{f^{(3)}(t)}{2!}(x-t)^2(-1)+\frac{f^{(4)}(t)}{3!}(x-t)^3\right)+\cdots+\\
&+\left(\frac{f^{(N)}(t)}{(N-1)!}(x-t)^{N-1}(-1)+\frac{f^{(N+1)}(t)}{N!}(x-t)^N\right)\\
&+B(N+1)(x-t)^N(-1).
\end{aligned}
$$

The second term in each parenthesis cancel with the first term in the next one, leaving just

$$F'(t)=\frac{f^{(N+1)}(t)}{N!}(x-t)^N+B(N+1)(x-t)^N(-1).$$

At some z, $F'(z)=0$ so

$$
\begin{aligned}
0&=\frac{f^{(N+1)}(z)}{N!}(x-z)^N+B(N+1)(x-z)^N(-1)\\
B(N+1)(x-z)^N&=\frac{f^{(N+1)}(z)}{N!}(x-z)^N\\
B&=\frac{f^{(N+1)}(z)}{(N+1)!}.
\end{aligned}
$$

Now we can write

$$F(t)=\sum_{n=0}^{N}\frac{f^{(n)}(t)}{n!}(x-t)^n+\frac{f^{(N+1)}(z)}{(N+1)!}(x-t)^{N+1}.$$

Recalling that $F(a)=f(x)$ we get

$$f(x)=F(a)=\sum_{n=0}^{N}\frac{f^{(n)}(a)}{n!}(x-a)^n+\frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1},$$

which is what we wanted to show. ♣

It may not be immediately obvious that this is particularly useful; let's look at some examples.

Example 9.57:

Find a polynomial approximation for $\sin x$ *accurate to* ± 0.005 *for values of* x *in* $[-\pi/2,\pi/2]$.

Solution. From Taylor's theorem with $a=0$:

$$\sin x=\sum_{n=0}^{N}\frac{f^{(n)}(a)}{n!}x^n+\frac{f^{(N+1)}(z)}{(N+1)!}x^{N+1}.$$

What can we say about the size of the term

$$\frac{f^{(N+1)}(z)}{(N+1)!}x^{N+1}?$$

Every derivative of $\sin x$ is $\pm\sin x$ or $\pm\cos x$, so $|f^{(N+1)}(z)| \leq 1$.

So we need to pick N so that

$$\left|\frac{x^{N+1}}{(N+1)!}\right| < 0.005.$$

Since we have limited x to $[-\pi/2, \pi/2]$,

$$\left|\frac{x^{N+1}}{(N+1)!}\right| < \frac{2^{N+1}}{(N+1)!}.$$

The quantity on the right decreases with increasing N, so all we need to do is find an N so that

$$\frac{2^{N+1}}{(N+1)!} < 0.005.$$

A little trial and error shows that $N = 8$ works, and in fact $2^9/9! < 0.0015$, so

$$\begin{aligned}\sin x &= \sum_{n=0}^{8} \frac{f^{(n)}(0)}{n!} x^n \pm 0.0015 \\ &= x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} \pm 0.0015.\end{aligned}$$

Figure 9.5 shows the graphs of $\sin x$ and and the approximation on $[0, 3\pi/2]$. As x gets larger, the approximation heads to negative infinity very quickly, since it is essentially acting like $-x^7$. ♣

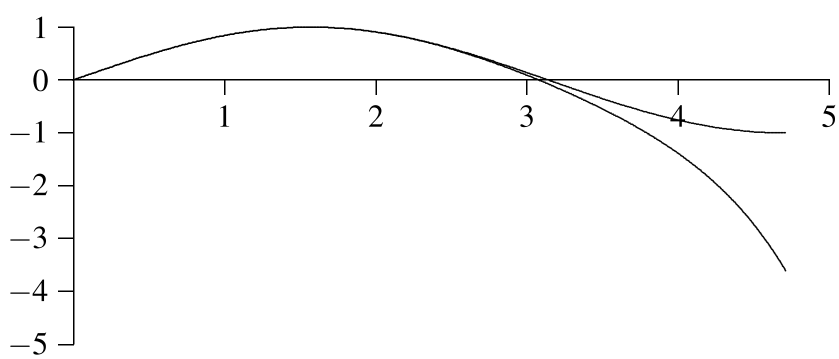


Figure 9.5: $\sin x$ **and a polynomial approximation.**

Note that we can now approximate the value of $\sin(x)$ to within 0.005 by using simple trigonometric identities to translate x into the interval $[-\pi/2, \pi/2]$.

We can extract a bit more information from this example. If we do not limit the value of x, we still have

$$\left|\frac{f^{(N+1)}(z)}{(N+1)!} x^{N+1}\right| \leq \left|\frac{x^{N+1}}{(N+1)!}\right|$$

so that $\sin x$ is represented by

$$\sum_{n=0}^{N} \frac{f^{(n)}(0)}{n!} x^n \pm \left|\frac{x^{N+1}}{(N+1)!}\right|.$$

If we can show that

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

for each x then

$$\sin x = \sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n = \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!},$$

that is, the sine function is actually equal to its Maclaurin series for all x. How can we prove that the limit is zero? Suppose that N is larger than $|x|$, and let M be the largest integer less than $|x|$ (if $M = 0$ the following is even easier). Then

$$\begin{aligned}\frac{|x^{N+1}|}{(N+1)!} &= \frac{|x|}{N+1}\frac{|x|}{N}\frac{|x|}{N-1}\cdots\frac{|x|}{M+1}\frac{|x|}{M}\frac{|x|}{M-1}\cdots\frac{|x|}{2}\frac{|x|}{1}\\ &\le \frac{|x|}{N+1}\cdot 1\cdot 1\cdots 1\cdot\frac{|x|}{M}\frac{|x|}{M-1}\cdots\frac{|x|}{2}\frac{|x|}{1}\\ &= \frac{|x|}{N+1}\frac{|x|^M}{M!}.\end{aligned}$$

The quantity $|x|^M/M!$ is a constant, so

$$\lim_{N\to\infty}\frac{|x|}{N+1}\frac{|x|^M}{M!} = 0$$

and by the Squeeze Theorem (9.6)

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

as desired. Essentially the same argument works for $\cos x$ and e^x; unfortunately, it is more difficult to show that most functions are equal to their Maclaurin series.

Example 9.58:

Find a polynomial approximation for e^x near $x = 2$ accurate to ± 0.005.

Solution. From Taylor's theorem:

$$e^x = \sum_{n=0}^{N}\frac{e^2}{n!}(x-2)^n + \frac{e^z}{(N+1)!}(x-2)^{N+1},$$

since $f^{(n)}(x) = e^x$ for all n. We are interested in x near 2, and we need to keep $|(x-2)^{N+1}|$ in check, so we may as well specify that $|x-2| \le 1$, so $x \in [1,3]$. Also

$$\left|\frac{e^z}{(N+1)!}\right| \le \frac{e^3}{(N+1)!},$$

so we need to find an N that makes $e^3/(N+1)! \le 0.005$. This time $N = 5$ makes $e^3/(N+1)! < 0.0015$, so the approximating polynomial is

$$e^x = e^2 + e^2(x-2) + \frac{e^2}{2}(x-2)^2 + \frac{e^2}{6}(x-2)^3 + \frac{e^2}{24}(x-2)^4 + \frac{e^2}{120}(x-2)^5 \pm 0.0015.$$

Note that our approximation requires that we already have a very accurate approximation of the value e^2, which we shouldn't assume we have in the context of trying to approximate e^x. For this reason we typically try to center our series on values for which the derivative of the function is easy to evaluate (e.g. $a = 0$). ♣

Note well that in these examples we found polynomials of a certain accuracy only on a small interval, even though the series for $\sin x$ and e^x converge for all x; this is typical. To get the same accuracy on a larger interval would require more terms.

Exercises for 9.11

Exercise 9.11.1 *Find a polynomial approximation for* $\cos x$ *on* $[0,\pi]$*, accurate to* $\pm 10^{-3}$

Exercise 9.11.2 *How many terms of the series for* $\ln x$ *centered at 1 are required so that the guaranteed error on* $[1/2,3/2]$ *is at most* 10^{-3}*? What if the interval is instead* $[1,3/2]$*?*

Exercise 9.11.3 *Find the first three nonzero terms in the Taylor series for* $\tan x$ *on* $[-\pi/4,\pi/4]$*, and compute the guaranteed error term as given by Taylor's theorem. (You may want to use Sage or a similar aid.)*

Exercise 9.11.4 *Show that* $\cos x$ *is equal to its Taylor series for all* x *by showing that the limit of the error term is zero as* N *approaches infinity.*

Exercise 9.11.5 *Show that* e^x *is equal to its Taylor series for all* x *by showing that the limit of the error term is zero as* N *approaches infinity.*

10. Differential Equations

Many physical phenomena can be modeled using the language of calculus. For example, observational evidence suggests that the temperature of a cup of tea (or some other liquid) in a room of constant temperature will cool over time at a rate proportional to the difference between the room temperature and the temperature of the tea.

In symbols, if t is the time, M is the room temperature, and $f(t)$ is the temperature of the tea at time t then $f'(t) = k(M - f(t))$ where $k > 0$ is a constant which will depend on the kind of tea (or more generally the kind of liquid) but not on the room temperature or the temperature of the tea. This is **Newton's law of cooling** and the equation that we just wrote down is an example of a **differential equation**. Ideally we would like to solve this equation, namely, find the function $f(t)$ that describes the temperature over time, though this often turns out to be impossible, in which case various approximation techniques must be used. The use and solution of differential equations is an important field of mathematics; here we see how to solve some simple but useful types of differential equation.

Informally, a differential equation is an equation in which one or more of the derivatives of some function appears. Typically, a scientific theory will produce a differential equation (or a system of differential equations) that describes or governs some physical process, but the theory will not produce the desired function or functions directly.

10.1 First Order Differential Equations

We start by considering equations in which only the first derivative of the function appears.

Definition 10.1: First Order Differential Equation

A **first order differential equation** *is an equation of the form $F(t, y, y') = 0$. A solution of a first order differential equation is a function $f(t)$ that makes $F(t, f(t), f'(t)) = 0$ for every value of t.*

Here, F is a function of three variables which we label t, y, and y'. It is understood that y' will explicitly appear in the equation although t and y need not. The term "first order" means that the first derivative of y appears, but no higher order derivatives do.

Example 10.2: Newton's Law of Cooling

The equation from Newton's law of cooling, $y' = k(M - y)$ is a first order differential equation; $F(t, y, y') = k(M - y) - y'$.

Example 10.3: A First Order Differential Equation

$y' = t^2 + 1$ is a first order differential equation; $F(t, y, y') = y' - t^2 - 1$. All solutions to this equation are of the form $t^3/3 + t + C$.

Definition 10.4: First Order Initial Value Problem

*A **first order initial value problem** is a system of equations of the form $F(t, y, y') = 0$, $y(t_0) = y_0$. Here t_0 is a fixed time and y_0 is a number. A solution of an initial value problem is a solution $f(t)$ of the differential equation that also satisfies the **initial condition** $f(t_0) = y_0$.*

Example 10.5: An Initial Value Problem

Verify that the initial value problem $y' = t^2 + 1$, $y(1) = 4$ has solution $f(t) = t^3/3 + t + 8/3$.

Solution. Observe that $f'(t) = t^2 + 1$ and $f(1) = 1^3/2 + 1 + 8/3 = 4$ as required. ♣

The general first order equation is too general, so we can't describe methods that will work on them all, or even a large portion of them. We can make progress with specific kinds of first order differential equations. For example, much can be said about equations of the form $y' = \phi(t, y)$ where ϕ is a function of the two variables t and y. Under reasonable conditions on ϕ, such an equation has a solution and the corresponding initial value problem has a unique solution. However, in general, these equations can be very difficult or impossible to solve explicitly.

Example 10.6: IVP for Newton's Law of Cooling

Consider this specific example of an initial value problem for Newton's law of cooling: $y' = 2(25 - y)$, $y(0) = 40$. Discuss the solutions for this initial value problem.

Solution. We first note the zero of the equation: If $y(t_0) = 25$, the right hand side of the differential equation is zero, and so the constant function $y(t) = 25$ is a solution to the differential equation. It is not a solution to the initial value problem, since $y(0) \neq 25$. (The physical interpretation of this constant solution is that if a liquid is at the same temperature as its surroundings, then the liquid will stay at that temperature.) So long as y is not 25, we can rewrite the differential equation as

$$\begin{aligned} \frac{dy}{dt}\frac{1}{25-y} &= 2 \\ \frac{1}{25-y}\,dy &= 2\,dt, \end{aligned}$$

so

$$\int \frac{1}{25-y}\,dy = \int 2\,dt,$$

that is, the two anti-derivatives must be the same except for a constant difference. We can calculate these anti-derivatives and rearrange the results:

$$\int \frac{1}{25-y}\,dy = \int 2\,dt$$

$$\begin{aligned}
(-1)\ln|25-y| &= 2t+C_0\\
\ln|25-y| &= -2t-C_0 = -2t+C\\
|25-y| &= e^{-2t+C} = e^{-2t}e^{C}\\
y-25 &= \pm e^{C}e^{-2t}\\
y &= 25\pm e^{C}e^{-2t} = 25+Ae^{-2t}.
\end{aligned}$$

Here $A=\pm e^{C}=\pm e^{-C_0}$ is some non-zero constant. Since we want $y(0)=40$, we substitute and solve for A:

$$\begin{aligned}
40 &= 25+Ae^{0}\\
15 &= A,
\end{aligned}$$

and so $y=25+15e^{-2t}$ is a solution to the initial value problem. Note that y is never 25, so this makes sense for all values of t. However, if we allow $A=0$ we get the solution $y=25$ to the differential equation, which would be the solution to the initial value problem if we were to require $y(0)=25$. Thus, $y=25+Ae^{-2t}$ describes all solutions to the differential equation $y'=2(25-y)$, and all solutions to the associated initial value problems. ♣

Why could we solve this problem? Our solution depended on rewriting the equation so that all instances of y were on one side of the equation and all instances of t were on the other. Of course, in this case the only t was originally hidden, since we didn't write dy/dt in the original equation. This is not required, however.

Example 10.7: Solving an IVP

Solve the differential equation $y'=2t(25-y)$.

Solution. This is almost identical to the previous example. As before, $y(t)=25$ is a solution. If $y\neq 25$,

$$\begin{aligned}
\int \frac{1}{25-y}\,dy &= \int 2t\,dt\\
(-1)\ln|25-y| &= t^2+C_0\\
\ln|25-y| &= -t^2-C_0 = -t^2+C\\
|25-y| &= e^{-t^2+C} = e^{-t^2}e^{C}\\
y-25 &= \pm e^{C}e^{-t^2}\\
y &= 25\pm e^{C}e^{-t^2} = 25+Ae^{-t^2}.
\end{aligned}$$

As before, all solutions are represented by $y=25+Ae^{-t^2}$, allowing A to be zero. ♣

Definition 10.8: Separable Differential Equations

A first order differential equation is **separable** *if it can be written in the form*

$$y'=f(t)g(y).$$

As in the examples, we can attempt to solve a separable equation by converting to the form

$$\int \frac{1}{g(y)}\,dy = \int f(t)\,dt.$$

This technique is called **separation of variables**. The simplest (in principle) sort of separable equation is one in which $g(y) = 1$, in which case we attempt to solve

$$\int 1\,dy = \int f(t)\,dt.$$

We can do this if we can find an anti-derivative of $f(t)$.

As we have seen so far, a differential equation typically has an infinite number of solutions. Such a solution is called a **general solution**. A corresponding initial value problem will give rise to just one solution. Such a solution in which there are no unknown constants remaining is called a **particular solution**.

The general approach to separable equations is as follows: Suppose we wish to solve $y' = f(t)g(y)$ where f and g are continuous functions. If $g(a) = 0$ for some a then $y(t) = a$ is a constant solution of the equation, since in this case $y' = 0 = f(t)g(a)$. For example, $y' = y^2 - 1$ has constant solutions $y(t) = 1$ and $y(t) = -1$.

To find the nonconstant solutions, we note that the function $1/g(y)$ is continuous where $g \neq 0$, so $1/g$ has an antiderivative G. Let F be an antiderivative of f. Now we write

$$G(y) = \int \frac{1}{g(y)}\,dy = \int f(t)\,dt = F(t) + C,$$

so $G(y) = F(t) + C$. Now we solve this equation for y.

Of course, there are a few places this ideal description could go wrong: We need to be able to find the antiderivatives G and F, and we need to solve the final equation for y. The upshot is that the solutions to the original differential equation are the constant solutions, if any, and all functions y that satisfy $G(y) = F(t) + C$.

Example 10.9: Population Growth and Radioactive Decay

Analyze the differential equation $y' = ky$.

Solution. When $k > 0$, this describes certain simple cases of population growth: It says that the change in the population y is proportional to the population. The underlying assumption is that each organism in the current population reproduces at a fixed rate, so the larger the population the more new organisms are produced. While this is too simple to model most real populations, it is useful in some cases over a limited time. When $k < 0$, the differential equation describes a quantity that decreases in proportion to the current value; this can be used to model radioactive decay.

The constant solution is $y(t) = 0$; of course this will not be the solution to any interesting initial value problem. For the non-constant solutions, we proceed much as before:

$$\int \frac{1}{y}\,dy = \int k\,dt$$

$$\begin{aligned}
\ln|y| &= kt + C \\
|y| &= e^{kt}e^{C} \\
y &= \pm e^{C}e^{kt} \\
y &= Ae^{kt}.
\end{aligned}$$

Again, if we allow $A = 0$ this includes the constant solution, and we can simply say that $y = Ae^{kt}$ is the general solution. With an initial value we can easily solve for A to get the solution of the initial value problem. In particular, if the initial value is given for time $t = 0$, $y(0) = y_0$, then $A = y_0$ and the solution is $y = y_0e^{kt}$. ♣

Exercises for 10.1

Exercise 10.1.1 *Which of the following equations are separable?*

(a) $y' = \sin(ty)$

(b) $y' = e^t e^y$

(c) $yy' = t$

(d) $y' = (t^3 - t)\arcsin(y)$

(e) $y' = t^2 \ln y + 4t^3 \ln y$

Exercise 10.1.2 *Solve* $y' = 1/(1+t^2)$.

Exercise 10.1.3 *Solve the initial value problem* $y' = t^n$ *with* $y(0) = 1$ *and* $n \geq 0$.

Exercise 10.1.4 *Solve* $y' = \ln t$.

Exercise 10.1.5 *Identify the constant solutions (if any) of* $y' = t\sin y$.

Exercise 10.1.6 *Identify the constant solutions (if any) of* $y' = te^y$.

Exercise 10.1.7 *Solve* $y' = t/y$.

Exercise 10.1.8 *Solve* $y' = y^2 - 1$.

Exercise 10.1.9 *Solve* $y' = t/(y^3 - 5)$. *You may leave your solution in implicit form: that is, you may stop once you have done the integration, without solving for y.*

Exercise 10.1.10 *Find a non-constant solution of the initial value problem* $y' = y^{1/3}$, $y(0) = 0$, *using separation of variables. Note that the constant function* $y(t) = 0$ *also solves the initial value problem. This shows that an initial value problem can have more than one solution.*

Exercise 10.1.11 *Solve the equation for Newton's law of cooling leaving M and k unknown.*

Exercise 10.1.12 *After 10 minutes in Jean-Luc's room, his tea has cooled to 40° Celsius from 100° Celsius. The room temperature is 25° Celsius. How much longer will it take to cool to 35°?*

Exercise 10.1.13 *Solve the* **logistic equation** $y' = ky(M-y)$. *(This is a somewhat more reasonable population model in most cases than the simpler $y' = ky$.) Sketch the graph of the solution to this equation when $M = 1000$, $k = 0.002$, $y(0) = 1$.*

Exercise 10.1.14 *Suppose that $y' = ky$, $y(0) = 2$, and $y'(0) = 3$. What is y?*

Exercise 10.1.15 *A radioactive substance obeys the equation $y' = ky$ where $k < 0$ and y is the mass of the substance at time t. Suppose that initially, the mass of the substance is $y(0) = M > 0$. At what time does half of the mass remain? (This is known as the half life. Note that the half life depends on k but not on M.)*

Exercise 10.1.16 *Bismuth-210 has a half life of five days. If there is initially 600 milligrams, how much is left after 6 days? When will there be only 2 milligrams left?*

Exercise 10.1.17 *The half life of carbon-14 is 5730 years. If one starts with 100 milligrams of carbon-14, how much is left after 6000 years? How long do we have to wait before there is less than 2 milligrams?*

Exercise 10.1.18 *A certain species of bacteria doubles its population (or its mass) every hour in the lab. The differential equation that models this phenomenon is $y' = ky$, where $k > 0$ and y is the population of bacteria at time t. What is y?*

Exercise 10.1.19 *If a certain microbe doubles its population every 4 hours and after 5 hours the total population has mass 500 grams, what was the initial mass?*

10.2 First Order Homogeneous Linear Equations

A simple, but important and useful, type of separable equation is the **first order homogeneous linear equation**:

Definition 10.10: First Order Homogeneous Linear Equation

A first order homogeneous linear differential equation is one of the form $y' + p(t)y = 0$ or equivalently $y' = -p(t)y$.

"Homogeneous" refers to the zero on the right side of the equation, provided that y' and y are on the left. "Linear" in this definition indicates that both y' and y appear independently and explicitly; we don't see y' or y to any power greater than 1, or multiplied by each other (i.e. $y'y$).

Example 10.11: Linear Examples

The equation $y' = 2t(25-y)$ *can be written* $y' + 2ty = 50t$. *This is linear, but not homogeneous. The equation* $y' = ky$, *or* $y' - ky = 0$ *is linear and homogeneous, with a particularly simple* $p(t) = -k$. *The equation* $y' + y^2 = 0$ *is homogeneous, but not linear.*

Since first order homogeneous linear equations are separable, we can solve them in the usual way:

$$\begin{aligned} y' &= -p(t)y \\ \int \frac{1}{y}\,dy &= \int -p(t)\,dt \\ \ln|y| &= P(t) + C \\ y &= \pm e^{P(t)} \\ y &= Ae^{P(t)}, \end{aligned}$$

where $P(t)$ is an anti-derivative of $-p(t)$. As in previous examples, if we allow $A = 0$ we get the constant solution $y = 0$.

Example 10.12: Solving an IVP

Solve the initial value problem

$$y' + y\cos t = 0,$$

subject to $y(0) = 1/2$ *and* $y(2) = 1/2$.

Solution. We start with

$$P(t) = \int -\cos t\,dt = -\sin t,$$

so the general solution to the differential equation is

$$y = Ae^{-\sin t}.$$

To compute A we substitute:

$$\frac{1}{2} = Ae^{-\sin 0} = A,$$

so the solutions is

$$y = \frac{1}{2}e^{-\sin t}.$$

For the second problem,

$$\begin{aligned} \frac{1}{2} &= Ae^{-\sin 2} \\ A &= \frac{1}{2}e^{\sin 2} \end{aligned}$$

so the solution is

$$y = \frac{1}{2}e^{\sin 2}e^{-\sin t}.$$

♣

Example 10.13:

Solve the initial value problem $ty' + 3y = 0$, $y(1) = 2$, assuming $t > 0$.

Solution. We write the equation in standard form: $y' + 3y/t = 0$. Then

$$P(t) = \int -\frac{3}{t}\,dt = -3\ln t$$

and

$$y = Ae^{-3\ln t} = At^{-3}.$$

Substituting to find A: $2 = A(1)^{-3} = A$, so the solution is $y = 2t^{-3}$. ♣

Exercises for 10.2

Find the general solution of each equation in the following exercises.

Exercise 10.2.1 $y' + 5y = 0$

Exercise 10.2.2 $y' - 2y = 0$

Exercise 10.2.3 $y' + \dfrac{y}{1+t^2} = 0$

Exercise 10.2.4 $y' + t^2 y = 0$

In the following exercises, solve the initial value problem.

Exercise 10.2.5 $y' + y = 0$, $y(0) = 4$

Exercise 10.2.6 $y' - 3y = 0$, $y(1) = -2$

Exercise 10.2.7 $y' + y\sin t = 0$, $y(\pi) = 1$

Exercise 10.2.8 $y' + ye^t = 0$, $y(0) = e$

Exercise 10.2.9 $y' + y\sqrt{1+t^4} = 0$, $y(0) = 0$

Exercise 10.2.10 $y' + y\cos(e^t) = 0$, $y(0) = 0$

Exercise 10.2.11 $ty' - 2y = 0$, $y(1) = 4$

Exercise 10.2.12 $t^2 y' + y = 0$, $y(1) = -2$, $t > 0$

Exercise 10.2.13 $t^3 y' = 2y$, $y(1) = 1$, $t > 0$

Exercise 10.2.14 $t^3 y' = 2y$, $y(1) = 0$, $t > 0$

Exercise 10.2.15 *A function $y(t)$ is a solution of $y' + ky = 0$. Suppose that $y(0) = 100$ and $y(2) = 4$. Find k and find $y(t)$.*

Exercise 10.2.16 *A function $y(t)$ is a solution of $y' + t^k y = 0$. Suppose that $y(0) = 1$ and $y(1) = e^{-13}$. Find k and find $y(t)$.*

Exercise 10.2.17 *A bacterial culture grows at a rate proportional to its population. If the population is one million at $t = 0$ and 1.5 million at $t = 1$ hour, find the population as a function of time.*

Exercise 10.2.18 *A radioactive element decays with a half-life of 6 years. If a mass of the element weighs ten pounds at $t = 0$, find the amount of the element at time t.*

10.3 First Order Linear Equations

As you might guess, a first order linear differential equation has the form $y' + p(t)y = f(t)$. Not only is this closely related in form to the first order homogeneous linear equation, we can use what we know about solving homogeneous equations to solve the general linear equation.

Suppose that $y_1(t)$ and $y_2(t)$ are solutions to $y' + p(t)y = f(t)$. Let $g(t) = y_1 - y_2$. Then

$$\begin{aligned} g'(t) + p(t)g(t) &= y_1' - y_2' + p(t)(y_1 - y_2) \\ &= (y_1' + p(t)y_1) - (y_2' + p(t)y_2) \\ &= f(t) - f(t) = 0. \end{aligned}$$

In other words, $g(t) = y_1 - y_2$ is a solution to the homogeneous equation $y' + p(t)y = 0$. Turning this around, any solution to the linear equation $y' + p(t)y = f(t)$, call it y_1, can be written as $y_2 + g(t)$, for some particular y_2 and some solution $g(t)$ of the homogeneous equation $y' + p(t)y = 0$. Since we already know how to find all solutions of the homogeneous equation, finding just one solution to the equation $y' + p(t)y = f(t)$ will give us all of them.

How might we find that one particular solution to $y' + p(t)y = f(t)$? Again, it turns out that what we already know helps. We know that the general solution to the homogeneous equation $y' + p(t)y = 0$ looks like $Ae^{P(t)}$. We now make an inspired guess: Consider the function $v(t)e^{P(t)}$, in which we have replaced the constant parameter A with the function $v(t)$. This technique is called **variation of parameters**. For convenience write this as $s(t) = v(t)h(t)$, where $h(t) = e^{P(t)}$ is a solution to the homogeneous equation. Now let's compute a bit with $s(t)$:

$$\begin{aligned} s'(t) + p(t)s(t) &= v(t)h'(t) + v'(t)h(t) + p(t)v(t)h(t) \\ &= v(t)(h'(t) + p(t)h(t)) + v'(t)h(t) \\ &= v'(t)h(t). \end{aligned}$$

The last equality is true because $h'(t) + p(t)h(t) = 0$. Since $h(t)$ is a solution to the homogeneous equation. We are hoping to find a function $s(t)$ so that $s'(t) + p(t)s(t) = f(t)$; we will have such a function if we can arrange to have $v'(t)h(t) = f(t)$, that is, $v'(t) = f(t)/h(t)$. But this is as easy (or hard) as finding an anti-derivative of $f(t)/h(t)$. Putting this all together, the general solution to $y' + p(t)y = f(t)$ is

$$v(t)h(t) + Ae^{P(t)} = v(t)e^{P(t)} + Ae^{P(t)}.$$

Example 10.14: Solving an IVP

Find the solution of the initial value problem $y' + 3y/t = t^2$, $y(1) = 1/2$.

Solution. First we find the general solution; since we are interested in a solution with a given condition at $t = 1$, we may assume $t > 0$. We start by solving the homogeneous equation as usual; call the solution g:

$$g = Ae^{-\int (3/t)\,dt} = Ae^{-3\ln t} = At^{-3}.$$

Then as in the discussion, $h(t) = t^{-3}$ and $v'(t) = t^2/t^{-3} = t^5$, so $v(t) = t^6/6$. We know that every solution to the equation looks like

$$v(t)t^{-3} + At^{-3} = \frac{t^6}{6}t^{-3} + At^{-3} = \frac{t^3}{6} + At^{-3}.$$

Finally we substitute to find A:

$$\begin{aligned}\frac{1}{2} &= \frac{(1)^3}{6}+A(1)^{-3}=\frac{1}{6}+A\\ A &= \frac{1}{2}-\frac{1}{6}=\frac{1}{3}.\end{aligned}$$

The solution is then

$$y=\frac{t^3}{6}+\frac{1}{3}t^{-3}.$$

♣

Another common method for solving such a differential equation is by means of an **integrating factor**. In the differential equation $y'+p(t)y=f(t)$, we note that if we multiply through by a function $I(t)$ to get $I(t)y'+I(t)p(t)y=I(t)f(t)$, the left hand side looks like it could be a derivative computed by the product rule:

$$\frac{d}{dt}(I(t)y)=I(t)y'+I'(t)y.$$

Now if we could choose $I(t)$ so that $I'(t)=I(t)p(t)$, this would be exactly the left hand side of the differential equation. But this is just a first order homogeneous linear equation, and we know a solution is $I(t)=e^{Q(t)}$, where $Q(t)=\int p\,dt$; note that $Q(t)=-P(t)$, where $P(t)$ appears in the variation of parameters method and $P'(t)=-p$. Now the modified differential equation is

$$\begin{aligned}e^{-P(t)}y'+e^{-P(t)}p(t)y &= e^{-P(t)}f(t)\\ \frac{d}{dt}(e^{-P(t)}y) &= e^{-P(t)}f(t).\end{aligned}$$

Integrating both sides gives

$$\begin{aligned}e^{-P(t)}y &= \int e^{-P(t)}f(t)\,dt\\ y &= e^{P(t)}\int e^{-P(t)}f(t)\,dt.\end{aligned}$$

If you look carefully, you will see that this is exactly the same solution we found by variation of parameters, because $e^{-P(t)}f(t)=f(t)/h(t)$.

Some people find it easier to remember how to use the integrating factor method, rather than variation of parameters. Since ultimately they require the same calculation, you should use whichever of the two methods appeals to you more strongly. Using this method, the solution of the previous example would look just a bit different: Starting with $y'+3y/t=t^2$, we recall that the integrating factor is $e^{\int 3/t}=e^{3\ln t}=t^3$. Then we multiply through by the integrating factor and solve:

$$\begin{aligned}t^3y'+t^33y/t &= t^3t^2\\ t^3y'+t^23y &= t^5\\ \frac{d}{dt}(t^3y) &= t^5\\ t^3y &= t^6/6\\ y &= t^3/6.\end{aligned}$$

This is the same answer, of course, and the problem is then finished just as before.

Exercises for 10.3

In the following exercises, find the general solution of the equation.

Exercise 10.3.1 $y' + 4y = 8$

Exercise 10.3.2 $y' - 2y = 6$

Exercise 10.3.3 $y' + ty = 5t$

Exercise 10.3.4 $y' + e^t y = -2e^t$

Exercise 10.3.5 $y' - y = t^2$

Exercise 10.3.6 $2y' + y = t$

Exercise 10.3.7 $ty' - 2y = 1/t$, $t > 0$

Exercise 10.3.8 $ty' + y = \sqrt{t}$, $t > 0$

Exercise 10.3.9 $y' \cos t + y \sin t = 1$, $-\pi/2 < t < \pi/2$

Exercise 10.3.10 $y' + y \sec t = \tan t$, $-\pi/2 < t < \pi/2$

10.4 Approximation

We have seen how to solve a restricted collection of differential equations, or more accurately, how to attempt to solve them—we still may not be able to find the required anti-derivatives. Not surprisingly, non-linear equations can be even more difficult to solve. Yet much is known about solutions to some more general equations.

Suppose $\phi(t, y)$ is a function of two variables. A more general class of first order differential equations has the form $y' = \phi(t, y)$. This is not necessarily a linear first order equation, since ϕ may depend on y in some complicated way; note however that y' appears in a very simple form. Under suitable conditions on the function ϕ, it can be shown that every such differential equation has a solution, and moreover that for each initial condition the associated initial value problem has exactly one solution. In practical applications this is obviously a very desirable property.

Example 10.15: First Order Non-linear

The equation $y' = t - y^2$ is a first order non-linear equation, because y appears to the second power. We will not be able to solve this equation.

Example 10.16: Non-linear and Separable

The equation $y' = y^2$ is also non-linear, but it is separable and can be solved by separation of variables.

Not all differential equations that are important in practice can be solved exactly, so techniques have been developed to approximate solutions. We describe one such technique, **Euler's Method**, which is simple though not particularly useful compared to some more sophisticated techniques.

Suppose we wish to approximate a solution to the initial value problem $y' = \phi(t,y)$, $y(t_0) = y_0$, for $t \geq t_0$. Under reasonable conditions on ϕ, we know the solution exists, represented by a curve in the t-y plane; call this solution $f(t)$. The point (t_0, y_0) is of course on this curve. We also know the slope of the curve at this point, namely $\phi(t_0, y_0)$. If we follow the tangent line for a brief distance, we arrive at a point that should be almost on the graph of $f(t)$, namely $(t_0 + \Delta t, y_0 + \phi(t_0, y_0)\Delta t)$; call this point (t_1, y_1). Now we pretend, in effect, that this point really is on the graph of $f(t)$, in which case we again know the slope of the curve through (t_1, y_1), namely $\phi(t_1, y_1)$. So we can compute a new point, $(t_2, y_2) = (t_1 + \Delta t, y_1 + \phi(t_1, y_1)\Delta t)$ that is a little farther along, still close to the graph of $f(t)$ but probably not quite so close as (t_1, y_1). We can continue in this way, doing a sequence of straightforward calculations, until we have an approximation (t_n, y_n) for whatever time t_n we need. At each step we do essentially the same calculation, namely:

$$(t_{i+1}, y_{i+1}) = (t_i + \Delta t, y_i + \phi(t_i, y_i)\Delta t).$$

We expect that smaller time steps Δt will give better approximations, but of course it will require more work to compute to a specified time. It is possible to compute a guaranteed upper bound on how far off the approximation might be, that is, how far y_n is from $f(t_n)$. Suffice it to say that the bound is not particularly good and that there are other more complicated approximation techniques that do better.

Example 10.17: Approximating a Solution

Compute an approximation to the solution for $y' = t - y^2$, $y(0) = 0$, when $t = 1$.

Solution. We will use $\Delta t = 0.2$, which is easy to do even by hand, though we should not expect the resulting approximation to be very good. We get

$$\begin{aligned}
(t_1, y_1) &= (0 + 0.2, 0 + (0 - 0^2)0.2) = (0.2, 0) \\
(t_2, y_2) &= (0.2 + 0.2, 0 + (0.2 - 0^2)0.2) = (0.4, 0.04) \\
(t_3, y_3) &= (0.6, 0.04 + (0.4 - 0.04^2)0.2) = (0.6, 0.11968) \\
(t_4, y_4) &= (0.8, 0.11968 + (0.6 - 0.11968^2)0.2) = (0.8, 0.23681533952) \\
(t_5, y_5) &= (1.0, 0.23681533952 + (0.6 - 0.23681533952^2)0.2) = (1.0, 0.385599038513605)
\end{aligned}$$

So $y(1) \approx 0.3856$. As it turns out, this is not accurate to even one decimal place. Figure 10.1 shows these points connected by line segments (the lower curve) compared to a solution obtained by a much better approximation technique. Note that the shape is approximately correct even though the end points are quite far apart.

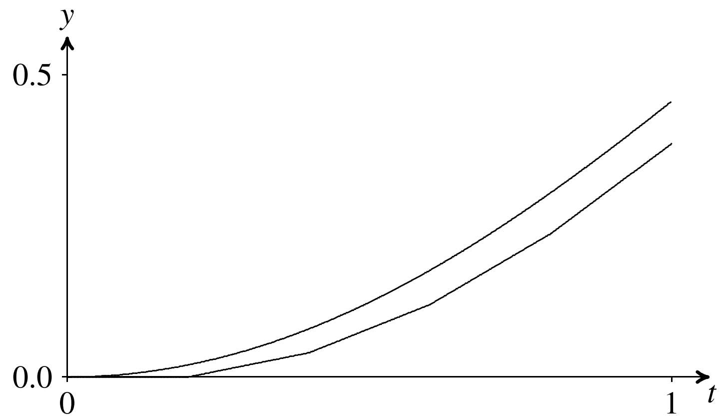


Figure 10.1: Approximating a solution to $y' = t - y^2$**,** $y(0) = 0$**.**

If you need to do Euler's method by hand, it is useful to construct a table to keep track of the work, as shown in Figure 10.2. Each row holds the computation for a single step: The starting point (t_i, y_i); the stepsize Δt; the computed slope $\phi(t_i, y_i)$; the change in y, $\Delta y = \phi(t_i, y_i)\Delta t$; and the new point, $(t_{i+1}, y_{i+1}) = (t_i + \Delta t, y_i + \Delta y)$. The starting point in each row is the newly computed point from the end of the previous row.

(t, y)	Δt	$\phi(t, y)$	$\Delta y = \phi(t, y)\Delta t$	$(t + \Delta t, y + \Delta y)$
$(0, 0)$	0.2	0	0	$(0.2, 0)$
$(0.2, 0)$	0.2	0.2	0.04	$(0.4, 0.04)$
$(0.4, 0.04)$	0.2	0.3984	0.07968	$(0.6, 0.11968)$
$(0.6, 0.11968)$	0.2	$0.58\ldots$	$0.117\ldots$	$(0.8, 0.236\ldots)$
$(0.8, 0.236\ldots)$	0.2	$0.743\ldots$	$0.148\ldots$	$(1.0, 0.385\ldots)$

Figure 10.2: Computing with Euler's Method.

♣

Euler's method is related to another technique that can help in understanding a differential equation in a qualitative way. Euler's method is based on the ability to compute the slope of a solution curve at any point in the plane, simply by computing $\phi(t, y)$. If we compute $\phi(t, y)$ at many points, say in a grid, and plot a small line segment with that slope at the point, we can get an idea of how solution curves must look. Such a plot is called a **slope field**. A slope field for $\phi = t - y^2$ is shown in Figure 10.3; compare this to figure 10.1. With a little practice, one can sketch reasonably accurate solution curves based on the slope field, in essence doing Euler's method visually.

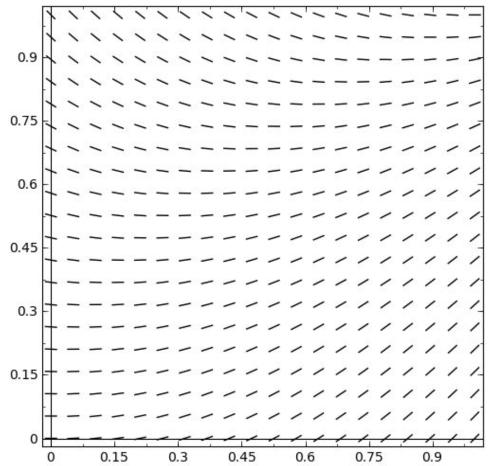


Figure 10.3: A slope field for $y' = t - y^2$.

Even when a differential equation can be solved explicitly, the slope field can help in understanding what the solutions look like with various initial conditions. Recall the logistic equation $y' = ky(M - y)$: y is a population at time t, M is a measure of how large a population the environment can support, and k measures the reproduction rate of the population. Figure 10.4 shows a slope field for this equation that is quite informative. It is apparent that if the initial population is smaller than M it rises to M over the long term, while if the initial population is greater than M it decreases to M.

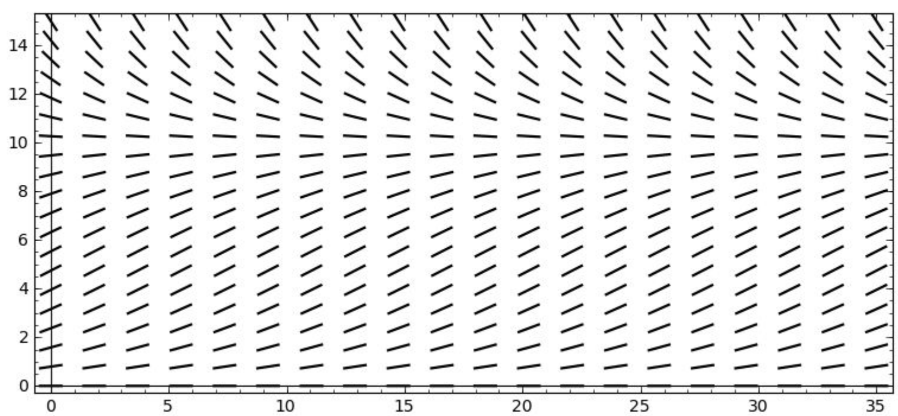


Figure 10.4: A slope field for $y' = 0.2y(10 - y)$.

Exercises for 10.4

In the following exercises, compute the Euler approximations for the initial value problem for $0 \leq t \leq 1$ and $\Delta t = 0.2$. If you have access to Sage, generate the slope field first and attempt to sketch the solution curve. Then use Sage to compute better approximations with smaller values of Δt.

Exercise 10.4.1 $y' = t/y$, $y(0) = 1$

Exercise 10.4.2 $y' = t + y^3$, $y(0) = 1$

Exercise 10.4.3 $y' = \cos(t + y)$, $y(0) = 1$

Exercise 10.4.4 $y' = t \ln y$, $y(0) = 2$

10.5 Second Order Homogeneous Equations

A second order differential equation is one containing the second derivative y''. These are in general quite complicated, but one fairly simple type is useful: The second order linear equation with constant coefficients.

Example 10.18: Second Order Homogeneous Equation

Analyze the intial value problem $y'' - y' - 2y = 0$, $y(0) = 5$, $y'(0) = 0$.

Solution. We make an inspired guess: might there be a solution of the form e^{rt}? This seems at least plausible, since in this case y'', y', and y all involve e^{rt}.

If such a function is a solution then

$$\begin{aligned} r^2 e^{rt} - re^{rt} - 2e^{rt} &= 0 \\ e^{rt}(r^2 - r - 2) &= 0 \\ (r^2 - r - 2) &= 0 \\ (r-2)(r+1) &= 0, \end{aligned}$$

so r is 2 or -1. Not only are $f = e^{2t}$ and $g = e^{-t}$ solutions, but notice that $y = Af + Bg$ is also, for any constants A and B:

$$\begin{aligned} (Af+Bg)'' - (Af+Bg)' - 2(Af+Bg) &= Af'' + Bg'' - Af' - Bg' - 2Af - 2Bg \\ &= A(f'' - f' - 2f) + B(g'' - g' - 2g) \\ &= A(0) + B(0) = 0. \end{aligned}$$

Can we find A and B so that this is a solution to the initial value problem? Let's substitute:

$$5 = y(0) = Af(0) + Bg(0) = Ae^0 + Be^0 = A + B$$

and

$$0 = y'(0) = Af'(0) + Bg'(0) = A2e^0 + B(-1)e^0 = 2A - B.$$

So we need to solve this system of <u>two</u> equations with <u>two</u> unknowns:

$$\begin{cases} A + B &= 5 \\ 2A - B &= 0 \end{cases}$$

Let $B = 2A$, substitute into the first equation to get $5 = A + 2A = 3A$. Then $A = 5/3$ and $B = 10/3$, and the desired solution is $(5/3)e^{2t} + (10/3)e^{-t}$. You now see why the initial condition in this case included both $y(0)$ and $y'(0)$: We needed two equations in the two unknowns A and B ♣

You should of course wonder whether there might be other solutions, but as it turns out, the answer is no. We will not prove this, but here is the theorem that tells us what we need to know:

Theorem 10.19: Solutions to Second Order Homogeneous

Given the differential equation $ay'' + by' + cy = 0$, $a \neq 0$, *consider the quadratic polynomial* $ar^2 + br + c = 0$, *called the* **characteristic polynomial**. *Using the quadratic formula, this polynomial always has one or two roots, call them* r_1 *and* r_2. *The general solution of the differential equation is:*

(a) $y = Ae^{r_1 t} + Be^{r_2 t}$, *if the roots* r_1 *and* r_2 *are real numbers, and* $r_1 \neq r_2$.

(b) $y = Ae^{r_1 t} + Bte^{r_2 t}$, *if* $r_1 = r_2$ *is a real, repeated root.*

(c) $y = A\cos(\beta t)e^{\alpha t} + B\sin(\beta t)e^{\alpha t}$, *if the roots are complex numbers,* $r = \alpha \pm \beta i$.

Example 10.20:

Solve the differential equation $y'' + A^2 y = 0$.

Solution. First we write the characteristic equation, $r^2 + A^2 = 0$. Then we find the roots of the characteristic equation:

$$r^2 = -A^2 \implies r = \pm Ai$$

These are imaginary roots, so the solution of the differential equation is in the form:

$$y = c_1 \cos(At) + c_2 \sin(At)$$

♣

Example 10.21:

Solve the differential equation $y'' - A^2 y = 0$.

Solution. First we write the characteristic equation, $r^2 - A^2 = 0$. Then we find the roots of the characteristic equation:

$$r^2 = A^2 \implies r = \pm A$$

These are imaginary roots, so the solution of the differential equation is in the form:

$$y = c_1 e^{At} + c_2 e^{-At} = c_1 e^{At} + \frac{c_2}{e^{At}}$$

♣

Example 10.22: Damped Spring Oscillation

Use a differential equation to describe the position of a mass hung on a spring.

Solution. Suppose a mass m is hung on a spring with spring constant k. If the spring is compressed or stretched and then released, the mass will oscillate up and down. Due to friction, the oscillation will be damped: Eventually the motion will cease. The damping will depend on the amount of friction; for example, if the system is suspended in oil the motion will cease sooner than if the system is in air. Using some simple physics, it is not hard to see that the position of the mass is described by the differential equation: $my'' + by' + ky = 0$. Using $m = 1$, $b = 4$, and $k = 5$ we find the motion of the mass. The characteristic polynomial is $r^2 + 4r + 5 = 0$, with roots $r = (-4 \pm \sqrt{16-20})/2 = -2 \pm i$. Thus the general solution is $y = A\cos(t)e^{-2t} + B\sin(t)e^{-2t}$. Suppose we know that $y(0) = 1$ and $y'(0) = 2$. Then as before we form two simultaneous equations: From $y(0) = 1$ we get $1 = A\cos(0)e^0 + B\sin(0)e^0 = A$. For the second we compute

$$y'' = -2Ae^{-2t}\cos(t) + Ae^{-2t}(-\sin(t)) - 2Be^{-2t}\sin(t) + Be^{-2t}\cos(t),$$

and then

$$2 = -2Ae^0\cos(0) - Ae^0\sin(0) - 2Be^0\sin(0) + Be^0\cos(0) = -2A + B.$$

So we get $A = 1$, $B = 4$, and $y = \cos(t)e^{-2t} + 4\sin(t)e^{-2t}$.

Here is a useful trick that makes this easier to understand: We have $y = (\cos t + 4\sin t)e^{-2t}$. The expression $\cos t + 4\sin t$ is a bit reminiscent of the trigonometric formula $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$ with $\alpha = t$. Let's rewrite it a bit as

$$\sqrt{17}\left(\frac{1}{\sqrt{17}}\cos t + \frac{4}{\sqrt{17}}\sin t\right).$$

Note that $(1/\sqrt{17})^2 + (4/\sqrt{17})^2 = 1$, which means that there is an angle β with $\cos\beta = 1/\sqrt{17}$ and $\sin\beta = 4/\sqrt{17}$ (of course, β may not be a "nice" angle). Then

$$\cos t + 4\sin t = \sqrt{17}(\cos t\cos\beta + \sin\beta\sin t) = \sqrt{17}\cos(t - \beta).$$

Thus, the solution may also be written $y = \sqrt{17}e^{-2t}\cos(t - \beta)$. This is a cosine curve that has been shifted β to the right; the $\sqrt{17}e^{-2t}$ has the effect of diminishing the amplitude of the cosine as t increases. ♣

Other physical systems that oscillate can also be described by such differential equations. Some electric circuits, for example, generate oscillating current.

Example 10.23:

Find the solution to the intial value problem $y'' - 4y' + 4y = 0$, $y(0) = -3$, $y'(0) = 1$.

Solution. The characteristic polynomial is $r^2 - 4r + 4 = (r-2)^2$, so there is one root, $r = 2$, and the general solution is $Ae^{2t} + Bte^{2t}$. Substituting $t = 0$ we get $-3 = A + 0 = A$. The first derivative is $2Ae^{2t} + 2Bte^{2t} + Be^{2t}$; substituting $t = 0$ gives $1 = 2A + 0 + B = 2A + B = 2(-3) + B = -6 + B$, so $B = 7$. The solution is $-3e^{2t} + 7te^{2t}$. ♣

Exercises for 10.5

Exercise 10.5.1 *Solve the initial value problem* $y'' - \omega^2 y = 0$, $y(0) = 1$, $y'(0) = 1$, *assuming* $\omega \neq 0$.

Exercise 10.5.2 *Solve the initial value problem* $2y'' + 18y = 0$, $y(0) = 2$, $y'(0) = 15$.

Exercise 10.5.3 *Solve the initial value problem* $y'' + 6y' + 5y = 0$, $y(0) = 1$, $y'(0) = 0$.

Exercise 10.5.4 *Solve the initial value problem* $y'' - y' - 12y = 0$, $y(0) = 0$, $y'(0) = 14$.

Exercise 10.5.5 *Solve the initial value problem* $y'' + 12y' + 36y = 0$, $y(0) = 5$, $y'(0) = -10$.

Exercise 10.5.6 *Solve the initial value problem* $y'' - 8y' + 16y = 0$, $y(0) = -3$, $y'(0) = 4$.

Exercise 10.5.7 *Solve the initial value problem* $y'' + 5y = 0$, $y(0) = -2$, $y'(0) = 5$.

Exercise 10.5.8 *Solve the initial value problem* $y'' + y = 0$, $y(\pi/4) = 0$, $y'(\pi/4) = 2$.

Exercise 10.5.9 *Solve the initial value problem* $y'' + 12y' + 37y = 0$, $y(0) = 4$, $y'(0) = 0$.

Exercise 10.5.10 *Solve the initial value problem* $y'' + 6y' + 18y = 0$, $y(0) = 0$, $y'(0) = 6$.

Exercise 10.5.11 *Solve the initial value problem* $y'' + 4y = 0$, $y(0) = \sqrt{3}$, $y'(0) = 2$.

Exercise 10.5.12 *Solve the initial value problem* $y'' + 100y = 0$, $y(0) = 5$, $y'(0) = 50$.

Exercise 10.5.13 *Solve the initial value problem* $y'' + 4y' + 13y = 0$, $y(0) = 1$, $y'(0) = 1$.

Exercise 10.5.14 *Solve the initial value problem* $y'' - 8y' + 25y = 0$, $y(0) = 3$, $y'(0) = 0$.

Exercise 10.5.15 *A mass-spring system* $my'' + by' + kx$ *has* $k = 29$, $b = 4$, *and* $m = 1$. *At time* $t = 0$ *the position is* $y(0) = 2$ *and the velocity is* $y'(0) = 1$. *Find* $y(t)$.

Exercise 10.5.16 *A mass-spring system* $my'' + by' + kx$ *has* $k = 24$, $b = 12$, *and* $m = 3$. *At time* $t = 0$ *the position is* $y(0) = 0$ *and the velocity is* $y'(0) = -1$. *Find* $y(t)$.

Exercise 10.5.17 *Consider the differential equation* $ay'' + by' = 0$, *with* a *and* b *both non-zero. Find the general solution by the method of this section. Now let* $g = y'$; *the equation may be written as* $ag' + bg = 0$, *a first order linear homogeneous equation. Solve this for* g, *then use the relationship* $g = y'$ *to find* y.

Exercise 10.5.18 *Suppose that* $y(t)$ *is a solution to* $ay'' + by' + cy = 0$, $y(t_0) = 0$, $y'(t_0) = 0$. *Show that* $y(t) = 0$.

10.6 Second Order Linear Equations - Method of Undetermined Coefficients

Now we consider second order equations of the form $ay'' + by' + cy = f(t)$, with a, b, and c constant. Of course, if $a = 0$ this is really a first order equation, so we assume $a \neq 0$. Also, if $c = 0$ we can solve the related first order equation $ah' + bh = f(t)$, and then solve $h = y'$ for y. So we will only examine examples in which $c \neq 0$.

Suppose that $y_1(t)$ and $y_2(t)$ are solutions to $ay'' + by' + cy = f(t)$, and consider the function $h = y_1 - y_2$. We substitute this function into the left hand side of the differential equation and simplify:

$$a(y_1 - y_2)'' + b(y_1 - y_2)' + c(y_1 - y_2) = ay_1'' + by_1' + cy_1 - (ay_2'' + by_2' + cy_2) = f(t) - f(t) = 0.$$

So h is a solution to the homogeneous equation $ay'' + by' + cy = 0$. Since we know how to find all such h, then with just one particular solution y_2 we can express all possible solutions y_1, namely, $y_1 = h + y_2$, where now h is the general solution to the homogeneous equation. Of course, this is exactly how we approached the first order linear equation.

To make use of this observation we need a method to find a single solution y_2. This turns out to be somewhat more difficult than the first order case, but if $f(t)$ is of a certain simple form, we can find a solution using the **method of undetermined coefficients**, sometimes more whimsically called the **method of judicious guessing**.

Example 10.24: Second Order Linear Equation

Solve the differential equation $y'' - y' - 6y = 18t^2 + 5$.

Solution. The general solution of the homogeneous equation is $Ae^{3t} + Be^{-2t}$. We guess that a solution to the non-homogeneous equation might look like $f(t)$ itself, namely, a quadratic $y = at^2 + bt + c$. Substituting this guess into the differential equation we get

$$y'' - y' - 6y = 2a - (2at + b) - 6(at^2 + bt + c) = -6at^2 + (-2a - 6b)t + (2a - b - 6c).$$

We want this to equal $18t^2 + 5$, so we need

$$\begin{aligned} -6a &= 18 \\ -2a - 6b &= 0 \\ 2a - b - 6c &= 5 \end{aligned}$$

This is a system of three equations in three unknowns and is not hard to solve: $a = -3$, $b = 1$, $c = -2$. Thus the general solution to the differential equation is $Ae^{3t} + Be^{-2t} - 3t^2 + t - 2$. ♣

So the "judicious guess" is a function with the same form as $f(t)$ but with undetermined (or better, yet to be determined) coefficients. This works whenever $f(t)$ is a polynomial.

Example 10.25: Mass-Spring System with No Damping

Analyze the initial value problem $my'' + ky = -mg$, $y(0) = 2$, $y'(0) = 50$.

Solution. The left hand side represents a mass-spring system with no damping, i.e., $b = 0$. Unlike the homogeneous case, we now consider the force due to gravity, $-mg$, assuming the spring is vertical at the surface of the earth, so that $g = 980$. To be specific, let us take $m = 1$ and $k = 100$. The general solution to the homogeneous equation is $A\cos(10t) + B\sin(10t)$. For the solution to the non-homogeneous equation we guess simply a constant $y = a$, since $-mg = -980$ is a constant. Then $y'' + 100y = 100a$ so $a = -980/100 = -9.8$. The desired general solution is then $A\cos(10t) + B\sin(10t) - 9.8$. Substituting the initial conditions we get

$$\begin{aligned} 2 &= A - 9.8 \\ 50 &= 10B \end{aligned}$$

so $A = 11.8$ and $B = 5$ and the solution is $11.8\cos(10t) + 5\sin(10t) - 9.8$. ♣

More generally, this method can be used when a function similar to $f(t)$ has derivatives that are also similar to $f(t)$; in the examples so far, since $f(t)$ was a polynomial, so were its derivatives. The method will work if $f(t)$ has the form $p(t)e^{\alpha t}\cos(\beta t) + q(t)e^{\alpha t}\sin(\beta t)$, where $p(t)$ and $q(t)$ are polynomials; when $\alpha = \beta = 0$ this is simply $p(t)$, a polynomial. In the most general form it is not simple to describe the appropriate judicious guess; we content ourselves with some examples to illustrate the process.

Example 10.26: Solving a Second Order Linear Equation

Find the general solution to $y'' + 7y' + 10y = e^{3t}$.

Solution. The characteristic equation is $r^2 + 7r + 10 = (r+5)(r+2)$, so the solution to the homogeneous equation is $Ae^{-5t} + Be^{-2t}$. For a particular solution to the inhomogeneous equation we guess Ce^{3t}. Substituting we get

$$9Ce^{3t} + 21Ce^{3t} + 10Ce^{3t} = e^{3t}40C.$$

When $C = 1/40$ this is equal to $f(t) = e^{3t}$, so the solution is $Ae^{-5t} + Be^{-2t} + (1/40)e^{3t}$. ♣

Example 10.27: Solving a Second Order Linear Equation

Find the general solution to $y'' + 7y' + 10y = e^{-2t}$.

Solution. Following the last example we might guess Ce^{-2t}, but since this is a solution to the homogeneous equation it cannot work. Instead we guess Cte^{-2t}. Then

$$(-2Ce^{-2t} - 2Ce^{-2t} + 4Cte^{-2t}) + 7(Ce^{-2t} - 2Cte^{-2t}) + 10Cte^{-2t} = e^{-2t}(-3C).$$

Then $C = -1/3$ and the solution is $Ae^{-5t} + Be^{-2t} - (1/3)te^{-2t}$. ♣

In general, if $f(t) = e^{kt}$ and k is one of the roots of the characteristic equation, then we guess Cte^{kt} instead of Ce^{kt}. If k is the only root of the characteristic equation, then Cte^{kt} will also not work, so we must guess Ct^2e^{kt}.

Example 10.28: Solving a Second Order Linear Equation

Find the general solution to $y'' - 6y' + 9y = e^{3t}$.

Solution. The characteristic equation is $r^2 - 6r + 9 = (r-3)^2$, so the general solution to the homogeneous equation is $Ae^{3t} + Bte^{3t}$. Guessing Ct^2e^{3t} for the particular solution, we get

$$(9Ct^2e^{3t} + 6Cte^{3t} + 6Cte^{3t} + 2Ce^{3t}) - 6(3Ct^2e^{3t} + 2Cte^{3t}) + 9Ct^2e^{3t} = e^{3t}2C.$$

Thus, the solution is $Ae^{3t} + Bte^{3t} + (1/2)t^2e^{3t}$. ♣

It is common in various physical systems to encounter an $f(t)$ of the form $a\cos(\omega t) + b\sin(\omega t)$.

Example 10.29: Solving a Second Order Linear Equation

Find the general solution to $y'' + 6y' + 25y = \cos(4t)$.

Solution. The roots of the characteristic equation are $-3 \pm 4i$, so the solution to the homogeneous equation is $e^{-3t}(A\cos(4t) + B\sin(4t))$. For a particular solution, we guess $C\cos(4t) + D\sin(4t)$. Substituting as usual:

$$(-16C\cos(4t) + -16D\sin(4t)) + 6(-4C\sin(4t) + 4D\cos(4t)) + 25(C\cos(4t) + D\sin(4t))$$
$$= (24D + 9C)\cos(4t) + (-24C + 9D)\sin(4t).$$

To make this equal to $\cos(4t)$ we need

$$\begin{aligned} 24D + 9C &= 1 \\ 9D - 24C &= 0 \end{aligned}$$

which gives $C = 1/73$ and $D = 8/219$. The full solution is then $e^{-3t}(A\cos(4t) + B\sin(4t)) + (1/73)\cos(4t) + (8/219)\sin(4t)$.

The function $e^{-3t}(A\cos(4t) + B\sin(4t))$ is a damped oscillation as in example 10.25, while $(1/73)\cos(4t) + (8/219)\sin(4t)$ is a simple undamped oscillation. As t increases, the sum $e^{-3t}(A\cos(4t) + B\sin(4t))$ approaches zero, so the solution

$$e^{-3t}(A\cos(4t) + B\sin(4t)) + (1/73)\cos(4t) + (8/219)\sin(4t)$$

becomes more and more like the simple oscillation $(1/73)\cos(4t) + (8/219)\sin(4t)$—notice that the initial conditions don't matter to this long term behavior. The damped portion is called the **transient** [part of the] **solution**, and the simple oscillation is called the **steady state** [part of the] **solution**. A physical example is a mass-spring system. If the only force on the mass is due to the spring, then the behavior of the system is a damped oscillation. If in addition an external force is applied to the mass, and if the force varies according to a function of the form $a\cos(\omega t) + b\sin(\omega t)$, then the long term behavior will be a simple oscillation determined by the steady state portion of the general solution; the initial position of the mass will not matter. ♣

As with the exponential form, such a simple guess may not work.

Example 10.30: Solving a Second Order Linear Equation

Find the general solution to $y'' + 16y = -\sin(4t)$.

Solution. The roots of the characteristic equation are $\pm 4i$, so the solution to the homogeneous equation is $A\cos(4t) + B\sin(4t)$. Since both $\cos(4t)$ and $\sin(4t)$ are solutions to the homogeneous equation, $C\cos(4t) + D\sin(4t)$ is also, so it cannot be a solution to the non-homogeneous equation. Instead, we guess $Ct\cos(4t) + Dt\sin(4t)$. Then substituting:

$$(-16Ct\cos(4t) - 16D\sin(4t) + 8D\cos(4t) - 8C\sin(4t))) + 16(Ct\cos(4t) + Dt\sin(4t))$$
$$= 8D\cos(4t) - 8C\sin(4t).$$

Thus $C = 1/8$, $D = 0$, and the solution is $C\cos(4t) + D\sin(4t) + (1/8)t\cos(4t)$. ♣

In general, if $f(t) = a\cos(\omega t) + b\sin(\omega t)$, and $\pm\omega i$ are the roots of the characteristic equation, then instead of $C\cos(\omega t) + D\sin(\omega t)$ we guess $Ct\cos(\omega t) + Dt\sin(\omega t)$.

Exercises for 10.6

Find the general solution to the differential equation.

Exercise 10.6.1 $y'' - 10y' + 25y = \cos t$

Exercise 10.6.2 $y'' + 2\sqrt{2}y' + 2y = 10$

Exercise 10.6.3 $y'' + 16y = 8t^2 + 3t - 4$

Exercise 10.6.4 $y'' + 2y = \cos(5t) + \sin(5t)$

Exercise 10.6.5 $y'' - 2y' + 2y = e^{2t}$

Exercise 10.6.6 $y'' - 6y + 13 = 1 + 2t + e^{-t}$

Exercise 10.6.7 $y'' + y' - 6y = e^{-3t}$

Exercise 10.6.8 $y'' - 4y' + 3y = e^{3t}$

Exercise 10.6.9 $y'' + 16y = \cos(4t)$

Exercise 10.6.10 $y'' + 9y = 3\sin(3t)$

Exercise 10.6.11 $y'' + 12y' + 36y = 6e^{-6t}$

Exercise 10.6.12 $y'' - 8y' + 16y = -2e^{4t}$

Exercise 10.6.13 $y'' + 6y' + 5y = 4$

Exercise 10.6.14 $y'' - y' - 12y = t$

Exercise 10.6.15 $y'' + 5y = 8\sin(2t)$

Exercise 10.6.16 $y'' - 4y = 4e^{2t}$

Solve the initial value problem.

Exercise 10.6.17 $y'' - y = 3t + 5,\ y(0) = 0,\ y'(0) = 0$

Exercise 10.6.18 $y'' + 9y = 4t,\ y(0) = 0,\ y'(0) = 0$

Exercise 10.6.19 $y'' + 12y' + 37y = 10e^{-4t},\ y(0) = 4,\ y'(0) = 0$

Exercise 10.6.20 $y'' + 6y' + 18y = \cos t - \sin t,\ y(0) = 0,\ y'(0) = 2$

Exercise 10.6.21 *Find the solution for the mass-spring equation* $y'' + 4y' + 29y = 689\cos(2t)$*.*

Exercise 10.6.22 *Find the solution for the mass-spring equation* $3y'' + 12y' + 24y = 2\sin t$*.*

Exercise 10.6.23 *Consider the differential equation* $my'' + by' + ky = \cos(\omega t)$*, with* m*,* b*, and* k *all positive and* $b^2 < 2mk$*; this equation is a model for a damped mass-spring system with external driving force* $\cos(\omega t)$*. Show that the steady state part of the solution has amplitude*

$$\frac{1}{\sqrt{(k - m\omega^2)^2 + \omega^2 b^2}}.$$

Show that this amplitude is largest when $\omega = \dfrac{\sqrt{4mk - 2b^2}}{2m}$*. This is the* **resonant frequency** *of the system.*

10.7 Second Order Linear Equations - Variation of Parameters

The method of the last section works only when the function $f(t)$ in $ay'' + by' + cy = f(t)$ has a particularly nice form, namely, when the derivatives of f look much like f itself. In other cases we can try variation of parameters as we did in the first order case.

Since as before $a \neq 0$, we can always divide by a to make the coefficient of y'' equal to 1. Thus, to simplify the discussion, we assume $a = 1$. We know that the differential equation $y'' + by' + cy = 0$ has a general solution $y = Ay_1 + By_2$. As before, we guess a particular solution to $y'' + by' + cy = f(t)$; this time we use the guess $y = u(t)y_1 + v(t)y_2$. Compute the derivatives:

$$\begin{aligned} y' &= u'y_1 + uy_1' + v'y_2 + vy_2' \\ y'' &= u''y_1 + u'y_1' + u'y_1' + uy_1'' + v''y_2 + v'y_2' + v'y_2' + vy_2''. \end{aligned}$$

Now substituting:

$$y'' + by' + cy = u''y_1 + u'y_1' + u'y_1' + uy_1'' + v''y_2 + v'y_2' + v'y_2' + vy_2''$$

$$\begin{aligned}
&\quad +bu'y_1+buy_1'+bv'y_2+bvy_2'+cuy_1+cvy_2\\
&= (uy_1''+buy_1'+cuy_1)+(vy_2''+bvy_2'+cvy_2)\\
&\quad +b(u'y_1+v'y_2)+(u''y_1+u'y_1'+v''y_2+v'y_2')+(u'y_1'+v'y_2')\\
&= 0+0+b(u'y_1+v'y_2)+(u''y_1+u'y_1'+v''y_2+v'y_2')+(u'y_1'+v'y_2').
\end{aligned}$$

The first two terms in parentheses are zero because y_1 and y_2 are solutions to the associated homogeneous equation. Now we engage in some wishful thinking. If $u'y_1+v'y_2=0$, then we also have $u''y_1+u'y_1'+v''y_2+v'y_2'=0$ by taking derivatives of both sides. This reduces the entire expression to $u'y_1'+v'y_2'=0$. We want this to be $f(t)$, that is, we need $u'y_1'+v'y_2'=f(t)$. So we would very much like these equations to be true:

$$\begin{aligned}
u'y_1+v'y_2 &= 0\\
u'y_1'+v'y_2' &= f(t).
\end{aligned}$$

This is a system of two equations in the two unknowns u' and v', so we can solve as usual to get $u'=g(t)$ and $v'=h(t)$. Then we can find u and v by computing antiderivatives. This is of course the sticking point in the whole plan, since the antiderivatives may be impossible to find. Nevertheless, this sometimes works out and is worth a try.

Example 10.31: Variation of Parameters

Consider the equation $y''-5y'+6y=\sin t$. Solve using variation of parameters.

Solution. The solution to the homogeneous equation is $Ae^{2t}+Be^{3t}$, so the simultaneous equations to be solved are

$$\begin{aligned}
u'e^{2t}+v'e^{3t} &= 0\\
2u'e^{2t}+3v'e^{3t} &= \sin t.
\end{aligned}$$

If we multiply the first equation by 2 and subtract it from the second equation we get

$$\begin{aligned}
v'e^{3t} &= \sin t\\
v' &= e^{-3t}\sin t\\
v &= -\frac{1}{10}(3\sin t+\cos t)e^{-3t},
\end{aligned}$$

using integration by parts. Then from the first equation:

$$\begin{aligned}
u' &= -e^{-2t}v'e^{3t}=-e^{-2t}e^{-3t}\sin(t)e^{3t}=-e^{-2t}\sin t\\
u &= \frac{1}{5}(2\sin t+\cos t)e^{-2t}.
\end{aligned}$$

Now the particular solution we seek is

$$\begin{aligned}
ue^{2t}+ve^{3t} &= \frac{1}{5}(2\sin t+\cos t)e^{-2t}e^{2t}-\frac{1}{10}(3\sin t+\cos t)e^{-3t}e^{3t}\\
&= \frac{1}{5}(2\sin t+\cos t)-\frac{1}{10}(3\sin t+\cos t)\\
&= \frac{1}{10}(\sin t+\cos t),
\end{aligned}$$

and the solution to the differential equation is $Ae^{2t}+Be^{3t}+(\sin t+\cos t)/10$. For comparison (and practice) you might want to solve this using the method of undetermined coefficients—both techniques should yield the same result. ♣

Example 10.32: Variation of Parameters

The differential equation $y''-5y'+6y=e^t\sin t$ can be solved using the method of undetermined coefficients, though we have not seen any examples of such a solution. Again, we will solve it by variation of parameters.

Solution. The equations to be solved are

$$\begin{aligned} u'e^{2t}+v'e^{3t} &= 0 \\ 2u'e^{2t}+3v'e^{3t} &= e^t\sin t. \end{aligned}$$

If we multiply the first equation by 2 and subtract it from the second equation we get

$$\begin{aligned} v'e^{3t} &= e^t\sin t \\ v' &= e^{-3t}e^t\sin t = e^{-2t}\sin t \\ v &= -\frac{1}{5}(2\sin t+\cos t)e^{-2t}. \end{aligned}$$

Then substituting we get

$$\begin{aligned} u' &= -e^{-2t}v'e^{3t} = -e^{-2t}e^{-2t}\sin(t)e^{3t} = -e^{-t}\sin t \\ u &= \frac{1}{2}(\sin t+\cos t)e^{-t}. \end{aligned}$$

The particular solution is

$$\begin{aligned} ue^{2t}+ve^{3t} &= \frac{1}{2}(\sin t+\cos t)e^{-t}e^{2t}-\frac{1}{5}(2\sin t+\cos t)e^{-2t}e^{3t} \\ &= \frac{1}{2}(\sin t+\cos t)e^{t}-\frac{1}{5}(2\sin t+\cos t)e^{t} \\ &= \frac{1}{10}(\sin t+3\cos t)e^t, \end{aligned}$$

and the solution to the differential equation is $Ae^{2t}+Be^{3t}+e^t(\sin t+3\cos t)/10$. ♣

Example 10.33: Solving a DE

The differential equation $y''-2y'+y=e^t/t^2$ is not of the form amenable to the method of undetermined coefficients. Solve it using variation of parameters.

Solution. The solution to the homogeneous equation is Ae^t+Bte^t and so the simultaneous equations are

$$\begin{aligned} u'e^t+v'te^t &= 0 \\ u'e^t+v'te^t+v'e^t &= \frac{e^t}{t^2}. \end{aligned}$$

Subtracting the equations gives

$$\begin{aligned} v'e^t &= \frac{e^t}{t^2} \\ v' &= \frac{1}{t^2} \\ v &= -\frac{1}{t}. \end{aligned}$$

Then substituting we get

$$\begin{aligned} u'e^t &= -v'te^t = -\frac{1}{t^2}te^t \\ u' &= -\frac{1}{t} \\ u &= -\ln t. \end{aligned}$$

The solution is $Ae^t + Bte^t - e^t \ln t - e^t$. ♣

Exercises for 10.7

Find the general solution to the differential equation using variation of parameters.

Exercise 10.7.1 $y'' + y = \tan x$

Exercise 10.7.2 $y'' + y = e^{2t}$

Exercise 10.7.3 $y'' + 4y = \sec x$

Exercise 10.7.4 $y'' + 4y = \tan x$

Exercise 10.7.5 $y'' + y' - 6y = t^2 e^{2t}$

Exercise 10.7.6 $y'' - 2y' + 2y = e^t \tan(t)$

Exercise 10.7.7 $y'' - 2y' + 2y = \sin(t)\cos(t)$ *(This is rather messy when done by variation of parameters; compare to undetermined coefficients.)*

11. Polar Coordinates, Parametric Equations

11.1 Polar Coordinates

Coordinate systems are tools that let us use algebraic methods to understand geometry. While the **rectangular** (also called **Cartesian**) coordinates that we have been using are the most common, some problems are easier to analyze in alternate coordinate systems.

A coordinate system is a scheme that allows us to identify any point in the plane or in three-dimensional space by a set of numbers. In rectangular coordinates these numbers are interpreted, roughly speaking, as the lengths of the sides of a rectangle. In **polar coordinates** a point in the plane is identified by a pair of numbers (r,θ). The number θ measures the angle between the positive x-axis and a ray that goes through the point, as shown in Figure 12.15; the number r measures the distance from the origin to the point (the radius). Figure 12.15 shows the point with rectangular coordinates $(1,\sqrt{3})$ and polar coordinates $(2,\pi/3)$, 2 units from the origin and $\pi/3$ radians from the positive x-axis.

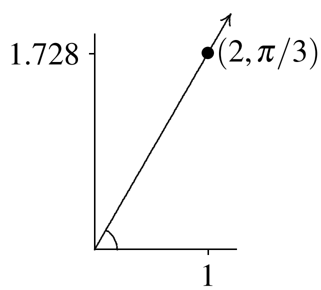


Figure 11.1: Polar coordinates of the point $(1,\sqrt{3})$**.**

Just as we describe curves in the plane using equations involving x and y, so can we describe curves using equations involving r and θ. Most common are equations of the form $r = f(\theta)$.

Example 11.1: Circle in Polar Coordinates

Graph the curve given by $r = 2$.

Solution. All points with $r = 2$ are at distance 2 from the origin, so $r = 2$ describes the circle of radius 2 with center at the origin. ♣

Example 11.2: Cardioid

Graph the curve given by $r = 1 + \cos\theta$.

Solution. We first consider $y = 1+\cos x$, as in figure 11.2. As θ goes through the values in $[0,2\pi]$, the value of r tracks the value of y, forming the "cardioid" shape of figure 11.2. For example, when $\theta = \pi/2$, $r = 1+\cos(\pi/2) = 1$, so we graph the point at distance 1 from the origin along the positive y-axis, which is at an angle of $\pi/2$ from the positive x-axis. When $\theta = 7\pi/4$, $r = 1+\cos(7\pi/4) = 1+\sqrt{2}/2 \approx 1.71$, and the corresponding point appears in the fourth quadrant. If we look at the curve, we should explain the indentation which occurs at $\theta = \pi$. It should be obvious that $\cos x < 0$ on $(\pi/2, 3\pi/2)$, thus $r = 1+\cos\theta < 1$. Specifically, when $\theta = \pi$, $\cos\pi = -1$, so $r = 0$, and we get the resulting indentation. This illustrates one of the potential benefits of using polar coordinates: the equation for this curve in rectangular coordinates would be *quite* complicated. ♣

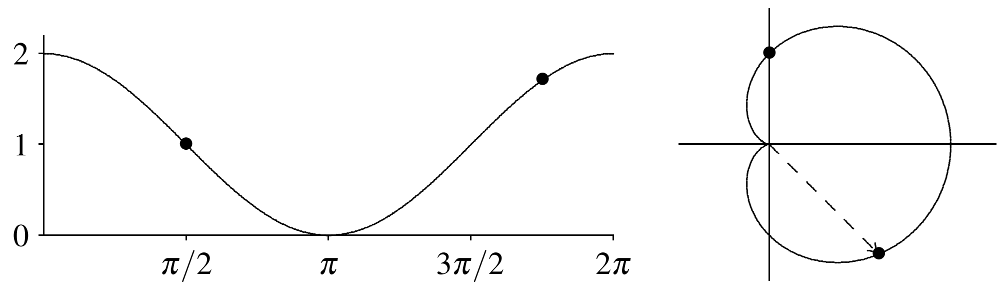


Figure 11.2: A cardioid: $y = 1+\cos x$ **on the left,** $r = 1+\cos\theta$ **on the right.**

Each point in the plane is associated with exactly one pair of numbers in the rectangular coordinate system; each point is associated with an infinite number of pairs in polar coordinates. In the cardioid example, we considered only the range $0 \le \theta \le 2\pi$, and already there was a duplicate: $(2,0)$ and $(2,2\pi)$ are the same point. Indeed, every value of θ outside the interval $[0,2\pi)$ duplicates a point on the curve $r = 1+\cos\theta$ when $0 \le \theta < 2\pi$. We can even make sense of polar coordinates like $(-2,\pi/4)$: Go to the direction $\pi/4$ and then move a distance 2 in the opposite direction; see figure 11.3. As usual, a negative angle θ means an angle measured clockwise from the positive x-axis. The point in figure 11.3 also has coordinates $(2,5\pi/4)$ and $(2,-3\pi/4)$.

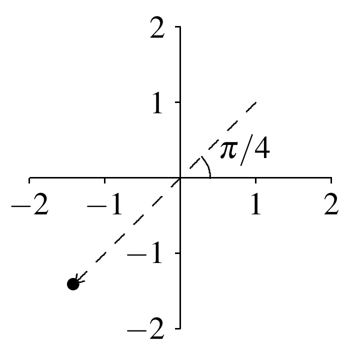


Figure 11.3: The point $(-2,\pi/4) = (2,5\pi/4) = (2,-3\pi/4)$ **in polar coordinates.**

The relationship between rectangular and polar coordinates is quite easy to understand. The point with polar coordinates (r,θ) has rectangular coordinates $x = r\cos\theta$ and $y = r\sin\theta$; this follows immediately from the definition of the sine and cosine functions. Using figure 11.3 as an example, the point shown has rectangular coordinates $x = (-2)\cos(\pi/4) = -\sqrt{2} \approx 1.4142$ and $y = (-2)\sin(\pi/4) = -\sqrt{2}$. This makes it very easy to convert equations from rectangular to polar coordinates.

Example 11.3: Straight Line in Polar Coordinates

Find the equation of the line $y = 3x + 2$ in polar coordinates.

Solution. We merely substitute: $r\sin\theta = 3r\cos\theta + 2$, or $r = \dfrac{2}{\sin\theta - 3\cos\theta}$. ♣

Example 11.4: Equation of a Circle

Find the equation of the circle $(x-1/2)^2 + y^2 = 1/4$ in polar coordinates.

Solution. Again substituting: $(r\cos\theta - 1/2)^2 + r^2\sin^2\theta = 1/4$. A bit of algebra turns this into $r = \cos(t)$. You should try plotting a few (r,θ) values to convince yourself that this makes sense. ♣

Example 11.5: Spiral of Archimedes

Graph the polar equation $r = \theta$.

Solution. Here the distance from the origin exactly matches the angle, so a bit of thought makes it clear that when $\theta \geq 0$ we get the spiral of Archimedes in figure 11.4. When $\theta < 0$, r is also negative, and so the full graph is the right hand picture in the figure. ♣

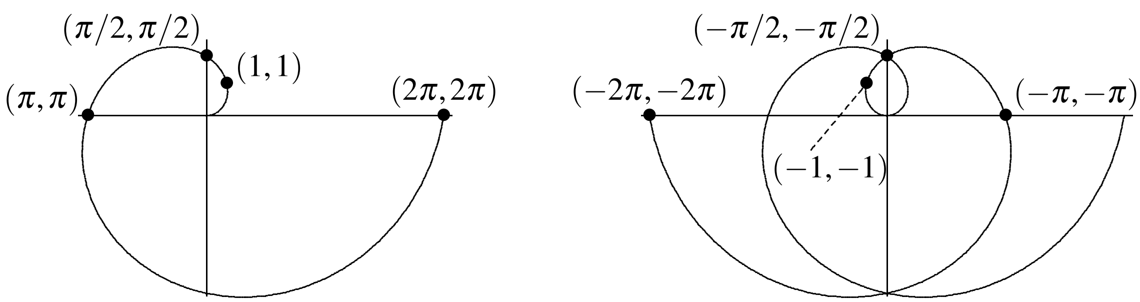


Figure 11.4: The spiral of Archimedes and the full graph of $r = \theta$.

Converting polar equations to rectangular equations can be somewhat trickier, and graphing polar equations directly is also not always easy.

Example 11.6: Graphing Polar Equations

Graph $r = 2\sin\theta$.

Solution. Given that the sine is periodic, we know that we will get the entire curve for values of θ in $[0, 2\pi)$. As θ runs from 0 to $\pi/2$, r increases from 0 to 2. Then as θ continues to π, r decreases again to 0. When θ runs from π to 2π, r is negative, and it is not hard to see that the first part of the curve is simply traced out again, so in fact we get the whole curve for values of θ in $[0, \pi)$. Thus, the curve looks something like figure 11.5. Now, this suggests that the curve could possibly be a circle, and if it is, it would have to be the circle $x^2+(y-1)^2=1$. Having made this guess, we can easily check it. First we substitute for x and y to get $(r\cos\theta)^2+(r\sin\theta-1)^2=1$; expanding and simplifying does indeed turn this into $r=2\sin\theta$. ♣

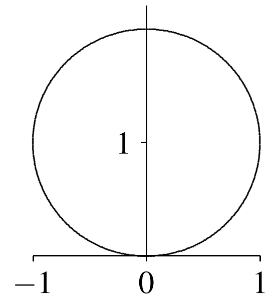


Figure 11.5: Graph of $r=2\sin\theta$.

Exercises for 11.1

Exercise 11.1.1 *Plot these polar coordinate points on one graph:* $(2,\pi/3)$, $(-3,\pi/2)$, $(-2,-\pi/4)$, $(1/2,\pi)$, $(1,4\pi/3)$, $(0,3\pi/2)$.

Exercise 11.1.2 *Find an equation in polar coordinates that has the same graph as the given equation in rectangular coordinates.*

(a) $y=3x$

(b) $y=-4$

(c) $xy^2=1$

(d) $x^2+y^2=5$

(e) $y=x^3$

(f) $y=\sin x$

(g) $y=5x+2$

(h) $x=2$

(i) $y=x^2+1$

Exercise 11.1.3 *Sketch the following curves:*

(a) $r=\cos\theta$

(b) $r=1+\theta^1/\pi^2$

(c) $r=\sin(\theta+\pi/4)$

(d) $r=-\sec\theta$

(e) $r=\theta/2,\ \theta\ge 0$

(f) $r=\cot\theta\csc\theta$

(g) $r=\dfrac{1}{\sin\theta+\cos\theta}$

(h) $r^2=-2\sec\theta\csc\theta$

Exercise 11.1.4 *Find an equation in rectangular coordinates that has the same graph as the given equation in polar coordinates.*

(a) $r = \sin(3\theta)$

(b) $r = \sin^2\theta$

(c) $r = \sec\theta\csc\theta$

(d) $r = \tan\theta$

11.2 Slopes in Polar Coordinates

When we describe a curve using polar coordinates, it is still a curve in the x-y plane. We would like to be able to compute slopes and areas for these curves using polar coordinates.

We have seen that $x = r\cos\theta$ and $y = r\sin\theta$ describe the relationship between polar and rectangular coordinates. If in turn we are interested in a curve given by $r = f(\theta)$, then we can write $x = f(\theta)\cos\theta$ and $y = f(\theta)\sin\theta$, describing x and y in terms of θ alone. The first of these equations describes θ implicitly in terms of x, so using the chain rule we may compute

$$\frac{dy}{dx} = \frac{dy}{d\theta}\frac{d\theta}{dx}.$$

Since $d\theta/dx = 1/(dx/d\theta)$, we can instead compute

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{f(\theta)\cos\theta + f'(\theta)\sin\theta}{-f(\theta)\sin\theta + f'(\theta)\cos\theta}.$$

Example 11.7: Horizontal Tangent Line

Find the points at which the curve given by $r = 1 + \cos\theta$ has a vertical or horizontal tangent line.

Solution. Since this function has period 2π, we may restrict our attention to the interval $[0, 2\pi)$ or $(-\pi, \pi]$, as convenience dictates. First, we compute the slope:

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

This fraction is zero when the numerator is zero (and the denominator is not zero). The numerator is $2\cos^2\theta + \cos\theta - 1$ so by the quadratic formula

$$\cos\theta = \frac{-1 \pm \sqrt{1+4\cdot 2}}{4} = -1 \quad \text{or} \quad \frac{1}{2}.$$

This means θ is π or $\pm\pi/3$. However, when $\theta = \pi$, the denominator is also 0, so we cannot conclude that the tangent line is horizontal.

Setting the denominator to zero we get

$$\begin{aligned} -\theta - 2\sin\theta\cos\theta &= 0 \\ \sin\theta(1 + 2\cos\theta) &= 0, \end{aligned}$$

so either $\sin\theta = 0$ or $\cos\theta = -1/2$. The first is true when θ is 0 or π, the second when θ is $2\pi/3$ or $4\pi/3$. However, as above, when $\theta = \pi$, the numerator is also 0, so we cannot conclude that the tangent line is vertical. Figure 11.6 shows points corresponding to θ equal to 0, ± 1.318, $2\pi/3$ and $4\pi/3$ on the graph of the function. Note that when $\theta = \pi$ the curve hits the origin and does not have a tangent line. ♣

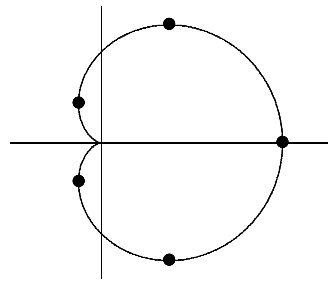


Figure 11.6: Points of vertical and horizontal tangency for $r = 1 + \cos\theta$.

We know that the second derivative $f''(x)$ is useful in describing functions, namely, in describing concavity. We can compute $f''(x)$ in terms of polar coordinates as well. We already know how to write $dy/dx = y'$ in terms of θ, then

$$\frac{d}{dx}\frac{dy}{dx} = \frac{dy'}{dx} = \frac{dy'}{d\theta}\frac{d\theta}{dx} = \frac{dy'/d\theta}{dx/d\theta}.$$

Example 11.8: Second Derivative of Cardioid

Find the second derivative for the cardioid $r = 1 + \cos\theta$.

Solution.

$$\frac{d}{d\theta}\frac{\cos\theta + \cos^2\theta - \sin^2\theta}{-\sin\theta - 2\sin\theta\cos\theta} \cdot \frac{1}{dx/d\theta} = \cdots = \frac{3(1+\cos\theta)}{(\sin\theta + 2\sin\theta\cos\theta)^2} \cdot \frac{1}{-(\sin\theta + 2\sin\theta\cos\theta)}$$
$$= \frac{-3(1+\cos\theta)}{(\sin\theta + 2\sin\theta\cos\theta)^3}.$$

The ellipsis here represents rather a substantial amount of algebra. We know from above that the cardioid has horizontal tangents at $\pm\pi/3$; substituting these values into the second derivative we get $y''(\pi/3) = -\sqrt{3}/2$ and $y''(-\pi/3) = \sqrt{3}/2$, indicating concave down and concave up respectively. This agrees with the graph of the function. ♣

Exercises for 11.2

Exercise 11.2.1 *Compute $y' = dy/dx$ and $y'' = d^2y/dx^2$.*

(a) $r = \theta$

(b) $r = 1 + \sin\theta$

(c) $r = \cos\theta$

(d) $r = \sin\theta$

(e) $r = \sec\theta$

(f) $r = \sin(2\theta)$

Exercise 11.2.2 *Sketch the curves over the interval $[0, 2\pi]$ unless otherwise stated.*

(a) $r = \sin\theta + \cos\theta$

(b) $r = 2 + 2\sin\theta$

(c) $r = \dfrac{3}{2} + \sin\theta$

(d) $r = 2 + \cos\theta$

(e) $r = \dfrac{1}{2} + \cos\theta$

(f) $r = \cos(\theta/2), 0 \le \theta \le 4\pi$

(g) $r = \sin(\theta/3), 0 \le \theta \le 6\pi$

(h) $r = \sin^2\theta$

(i) $r = 1 + \cos^2(2\theta)$

(j) $r = \sin^2(3\theta)$

(k) $r = \tan\theta$

(l) $r = \sec(\theta/2), 0 \le \theta \le 4\pi$

(m) $r = 1 + \sec\theta$

(n) $r = \dfrac{1}{1 - \cos\theta}$

(o) $r = \dfrac{1}{1 + \sin\theta}$

(p) $r = \cot(2\theta)$

(q) $r = \pi/\theta, 0 \le \theta \le \infty$

(r) $r = 1 + \pi/\theta, 0 \le \theta \le \infty$

11.3 Areas in Polar Coordinates

We can use the equation of a curve in polar coordinates to compute some areas bounded by such curves. The basic approach is the same as with any application of integration: Find an approximation that approaches the true value. For areas in rectangular coordinates, we approximated the region using rectangles; in polar coordinates, we use sectors of circles, as depicted in Figure 11.7. Recall that the area of a sector of a circle is $\alpha r^2/2$, where α is the angle subtended by the sector. If the curve is given by $r = f(\theta)$, and the angle subtended by a small sector is $\Delta\theta$, the area is $(\Delta\theta)(f(\theta))^2/2$. Thus we approximate the total area as

$$\sum_{i=0}^{n-1} \frac{1}{2} f(\theta_i)^2 \,\Delta\theta.$$

As a limit, this will give rise to:

$$\int_a^b \frac{1}{2} f(\theta)^2 \,d\theta.$$

Example 11.9: Area inside a Cardioid

Find the area inside the cardioid $r = 1 + \cos\theta$.

Solution.

$$\int_0^{2\pi} \frac{1}{2}(1+\cos\theta)^2 \,d\theta = \frac{1}{2}\int_0^{2\pi} 1 + 2\cos\theta + \cos^2\theta \,d\theta = \frac{1}{2}\left(\theta + 2\sin\theta + \frac{\theta}{2} + \frac{\sin 2\theta}{4}\right)\Bigg|_0^{2\pi} = \frac{3\pi}{2}.$$

♣

As in the example above, if a function is defined over the complete region $[0, 2\pi]$, then the bounds for the integral are $a = 0$, $b = 2\pi$. However, sometimes we need to be careful!

Example 11.10: Area inside a Loop

Find the area inside the loop $r = \sqrt{\cos\theta}$.

Solution. We look for points of intersection.

$$0 = \sqrt{\cos\theta}$$
$$0 = \cos\theta$$

This has solutions $\theta = \pi/2$ and $3\pi/2$, but we must select our bounds for integration correctly. Since $\cos\theta$ is negative on $(\pi/2, 3\pi/3)$, $\sqrt{\cos\theta}$ is undefined. Note that one full revolution is 2π; that is, the function is 2π periodic. If we were to integrate on $[3\pi/2, \pi/2]$ we would get the same answer, and can avoid a scenario in which $\sqrt{\cos\theta}$ is undefined. Equivalently, we can integrate on $[-\pi/2, \pi/2]$.

$$\int_a^b \frac{1}{2} f(\theta)^2 \, d\theta = \int_{-\pi/2}^{\pi/2} \frac{1}{2}\left(\sqrt{\cos\theta}\right)^2 d\theta = \frac{1}{2}\int_{-\pi/2}^{\pi/2} \cos\theta \, d\theta = \frac{1}{2}\left(\sin\theta\right)\Big|_{-\pi/2}^{\pi/2} = 1$$

♣

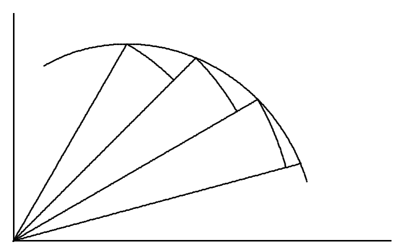


Figure 11.7: Approximating area by sectors of circles.

Example 11.11: Area Between Circles

Find the area between the circles $r = 2$ and $r = 4\sin\theta$, as shown in figure 11.8.

Solution. The two curves intersect where $2 = 4\sin\theta$, or $\sin\theta = 1/2$, so $\theta = \pi/6$ or $5\pi/6$. The area we want is then

$$\frac{1}{2}\int_{\pi/6}^{5\pi/6} 16\sin^2\theta - 4 \, d\theta = \frac{4}{3}\pi + 2\sqrt{3}.$$

♣

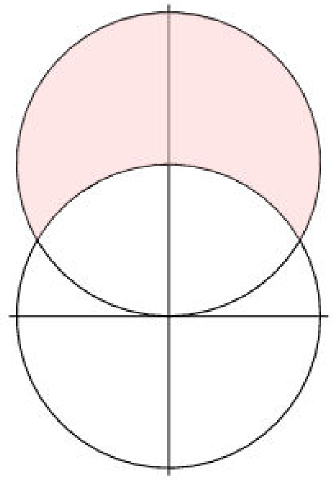


Figure 11.8: An area between curves.

This first example makes the process appear more straightforward than it is. Since points have many different representations in polar coordinates, it is not always so easy to identify points of intersection.

Example 11.12: Shaded Area

Find the shaded area in the first graph of figure 11.9 as the difference of the other two shaded areas. The cardioid is $r = 1 + \sin\theta$ and the circle is $r = 3\sin\theta$.

Solution. We attempt to find points of intersection:

$$\begin{aligned} 1 + \sin\theta &= 3\sin\theta \\ 1 &= 2\sin\theta \\ 1/2 &= \sin\theta. \end{aligned}$$

This has solutions $\theta = \pi/6$ and $5\pi/6$; $\pi/6$ corresponds to the intersection in the first quadrant that we need. Note that no solution of this equation corresponds to the intersection point at the origin, but fortunately that one is obvious. The cardioid goes through the origin when $\theta = -\pi/2$; the circle goes through the origin at multiples of π, starting with 0.

Now the larger region has area

$$\frac{1}{2}\int_{-\pi/2}^{\pi/6} (1 + \sin\theta)^2\, d\theta = \frac{\pi}{2} - \frac{9}{16}\sqrt{3},$$

and the smaller has area

$$\frac{1}{2}\int_{0}^{\pi/6} (3\sin\theta)^2\, d\theta = \frac{3\pi}{8} - \frac{9}{16}\sqrt{3},$$

so the difference is the area we seek, which is $\pi/8$. ♣

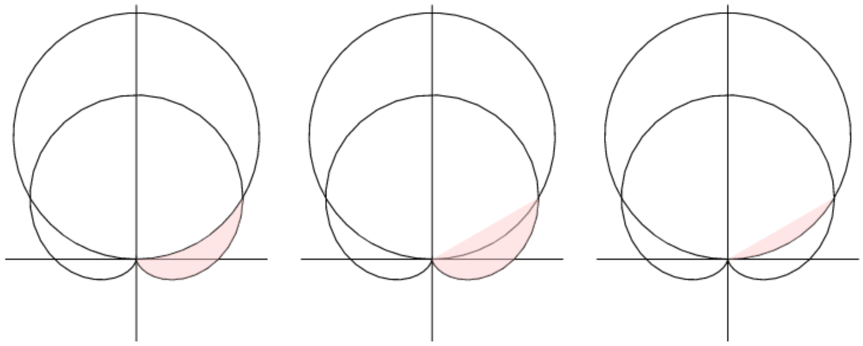


Figure 11.9: An area between curves.

Exercises for 11.3

Exercise 11.3.1 *Find the area enclosed by the curve.*

(a) $r = \sqrt{\sin\theta}$

(b) $r = 2 + \cos\theta$

(c) $r = \sec\theta, \pi/6 \le \theta \le \pi/3$

(d) $r = \cos\theta, 0 \le \theta \le \pi/3$

(e) $r = 2a\cos\theta, a > 0$

(f) $r = 4 + 3\sin\theta$

Exercise 11.3.2 *Find the area inside the loop formed by* $r = \tan(\theta/2)$.

Exercise 11.3.3 *Find the area inside one loop of* $r = \cos(3\theta)$.

Exercise 11.3.4 *Find the area inside one loop of* $r = \sin^2\theta$.

Exercise 11.3.5 *Find the area inside the small loop of* $r = (1/2) + \cos\theta$.

Exercise 11.3.6 *Find the area inside* $r = (1/2) + \cos\theta$*, including the area inside the small loop.*

Exercise 11.3.7 *Find the area inside one loop of* $r^2 = \cos(2\theta)$.

Exercise 11.3.8 *Find the area enclosed by* $r = \tan\theta$ *and* $r = \dfrac{\csc\theta}{\sqrt{2}}$.

Exercise 11.3.9 *Find the area inside* $r = 2\cos\theta$ *and outside* $r = 1$.

Exercise 11.3.10 *Find the area inside* $r = 2\sin\theta$ *and above the line* $r = (3/2)\csc\theta$.

Exercise 11.3.11 *Find the area inside* $r = \theta$, $0 \le \theta \le 2\pi$.

Exercise 11.3.12 *Find the area inside* $r = \sqrt{\theta}$, $0 \le \theta \le 2\pi$.

Exercise 11.3.13 *Find the area inside both* $r = \sqrt{3}\cos\theta$ *and* $r = \sin\theta$.

Exercise 11.3.14 *Find the area inside both* $r = 1 - \cos\theta$ *and* $r = \cos\theta$.

Exercise 11.3.15 *The center of a circle of radius 1 is on the circumference of a circle of radius 2. Find the area of the region inside both circles.*

Exercise 11.3.16 *Find the shaded area in figure 11.10. The curve is* $r = \theta$, $0 \le \theta \le 3\pi$.

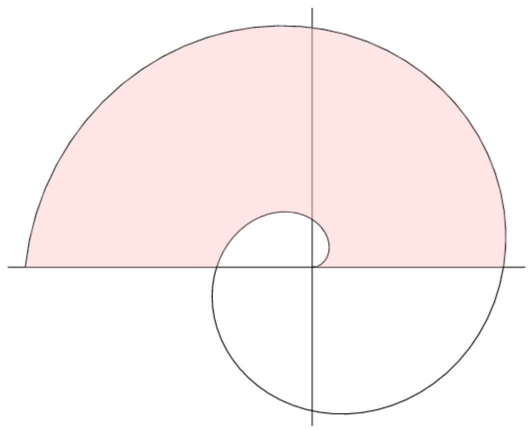


Figure 11.10: An area bounded by the spiral of Archimedes.

11.4 Parametric Equations

When we computed the derivative dy/dx using polar coordinates, we used the expressions $x = f(\theta)\cos\theta$ and $y = f(\theta)\sin\theta$. These two equations completely specify the curve, though the form $r = f(\theta)$ is simpler. The expanded form has the virtue that it can easily be generalized to describe a wider range of curves than can be specified in rectangular or polar coordinates.

Suppose $f(t)$ and $g(t)$ are functions. Then the equations $x = f(t)$ and $y = g(t)$ describe a curve in the plane. In the case of the polar coordinates equations, the variable t is replaced by θ which has a natural geometric interpretation. In general, t is simply an arbitrary variable, often called in this case a **parameter**, and this method of specifying a curve is known as **parametric equations**. One important interpretation of t is *time*. In this interpretation, the equations $x = f(t)$ and $y = g(t)$ give the position of an object at time t.

Example 11.13: Position of a Path

Describe the path of an object that moves so that its position at time t *is given by* $x = \cos t$, $y = \cos^2 t$.

Solution. We see immediately that $y = x^2$, so the path lies on this parabola. The path is not the entire parabola, however, since $x = \cos t$ is always between -1 and 1. It is now easy to see that the object oscillates back and forth on the parabola between the endpoints $(1,1)$ and $(-1,1)$, and is at point $(1,1)$ at time $t = 0$. ♣

It is sometimes quite easy to describe a complicated path in parametric equations when rectangular and polar coordinate expressions are difficult or impossible to devise.

Example 11.14: Wheel

A wheel of radius 1 rolls along a straight line, say the x-axis. A point on the rim of the wheel will trace out a curve, called a cycloid. Assume the point starts at the origin; find parametric equations for the curve.

Solution. Figure 11.11 illustrates the generation of the curve. The wheel is shown at its starting point, and again after it has rolled through about 490 degrees. We take as our parameter t the angle through which the wheel has turned, measured as shown clockwise from the line connecting the center of the wheel to the ground. Since the radius is 1, the center of the wheel has coordinates $(t,1)$. We seek to write the coordinates of the point on the rim as $(t+\Delta x, 1+\Delta y)$, where Δx and Δy are as shown in figure 11.12. These values are nearly the sine and cosine of the angle t, from the unit circle definition of sine and cosine. However, some care is required because we are measuring t from a nonstandard starting line and in a clockwise direction, as opposed to the usual counterclockwise direction. A bit of thought reveals that $\Delta x = -\sin t$ and $\Delta y = -\cos t$. Thus the parametric equations for the cycloid are $x = t - \sin t$, $y = 1 - \cos t$. ♣

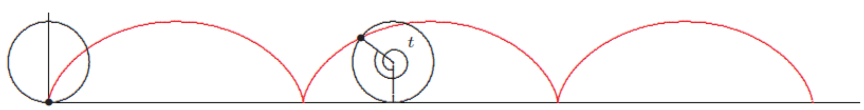


Figure 11.11: A cycloid.

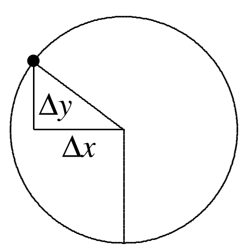


Figure 11.12: The wheel.

Exercises for 11.4

Exercise 11.4.1 *What curve is described by $x = t^2$, $y = t^4$? If t is interpreted as time, describe how the object moves on the curve.*

Exercise 11.4.2 *What curve is described by $x = 3\cos t$, $y = 3\sin t$? If t is interpreted as time, describe how the object moves on the curve.*

Exercise 11.4.3 *What curve is described by $x = 3\cos t$, $y = 2\sin t$? If t is interpreted as time, describe how the object moves on the curve.*

Exercise 11.4.4 *What curve is described by $x = 3\sin t$, $y = 3\cos t$? If t is interpreted as time, describe how the object moves on the curve.*

Exercise 11.4.5 *Sketch the curve described by $x = t^3 - t$, $y = t^2$. If t is interpreted as time, describe how the object moves on the curve.*

Exercise 11.4.6 *A wheel of radius 1 rolls along a straight line, say the x-axis. A point P is located halfway between the center of the wheel and the rim; assume P starts at the point $(0, 1/2)$. As the wheel rolls, P traces a curve; find parametric equations for the curve.*

11.5 Calculus with Parametric Equations

We have already seen how to compute slopes of curves given by parametric equations—it is how we computed slopes in polar coordinates.

Example 11.15: Slope of Cycloid

Find the slope of the cycloid $x = t - \sin t$, $y = 1 - \cos t$.

Solution. We compute $x' = 1 - \cos t$, $y' = \sin t$, so

$$\frac{dy}{dx} = \frac{\sin t}{1 - \cos t}.$$

Note that when t is an odd multiple of π, like π or 3π, this is $(0/2) = 0$, so there is a horizontal tangent line, in agreement with figure 11.11. At even multiples of π, the fraction is $0/0$, which is undefined. The figure shows that there is no tangent line at such points. ♣

Areas can be a bit trickier with parametric equations, depending on the curve and the area desired. We can potentially compute areas between the curve and the x-axis quite easily.

Example 11.16: Area Under Cycloid Arch

Find the area under one arch of the cycloid $x = t - \sin t$, $y = 1 - \cos t$.

Solution. We would like to compute

$$\int_0^{2\pi} y\,dx,$$

but we do not know y in terms of x. However, the parametric equations allow us to make a substitution: use $y = 1 - \cos t$ to replace y, and compute $dx = (1 - \cos t)\,dt$. Then the integral becomes

$$\int_0^{2\pi} (1 - \cos t)(1 - \cos t)\,dt = 3\pi.$$

Note that we need to convert the original x limits to t limits using $x = t - \sin t$. When $x = 0$, $t = \sin t$, which happens only when $t = 0$. Likewise, when $x = 2\pi$, $t - 2\pi = \sin t$ and $t = 2\pi$. Alternately, because we understand how the cycloid is produced, we can see directly that one arch is generated by $0 \le t \le 2\pi$. In general, of course, the t limits will be different than the x limits. ♣

This technique will allow us to compute some quite interesting areas, as illustrated by the exercises.

As a final example, we see how to compute the length of a curve given by parametric equations. The arc length for functions given as y in terms of x is the formula:

$$\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx.$$

Using some properties of derivatives, including the chain rule, we can convert this to use parametric equations $x = f(t)$, $y = g(t)$:

$$\begin{aligned}\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dx}{dt}\right)^2 \left(\frac{dy}{dx}\right)^2}\,\frac{dt}{dx}\,dx \\ &= \int_u^v \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt \\ &= \int_u^v \sqrt{(f'(t))^2 + (g'(t))^2}\,dt.\end{aligned}$$

Here u and v are the t limits corresponding to the x limits a and b.

Example 11.17: Length of Cycloid Arch

Find the length of one arch of the cycloid.

Solution. From $x = t - \sin t$, $y = 1 - \cos t$, we get the derivatives $f' = 1 - \cos t$ and $g' = \sin t$, so the length is

$$\int_0^{2\pi} \sqrt{(1 - \cos t)^2 + \sin^2 t}\,dt = \int_0^{2\pi} \sqrt{2 - 2\cos t}\,dt.$$

Now we use the formula $\sin^2(t/2) = (1-\cos(t))/2$ or $4\sin^2(t/2) = 2-2\cos t$ to get

$$\int_0^{2\pi} \sqrt{4\sin^2(t/2)}\,dt.$$

Since $0 \le t \le 2\pi$, $\sin(t/2) \ge 0$, so we can rewrite this as

$$\int_0^{2\pi} 2\sin(t/2)\,dt = 8.$$

♣

Exercises for 11.5

Exercise 11.5.1 *Consider the curve of 11.4.6 in section 11.4. Find all values of t for which the curve has a horizontal tangent line.*

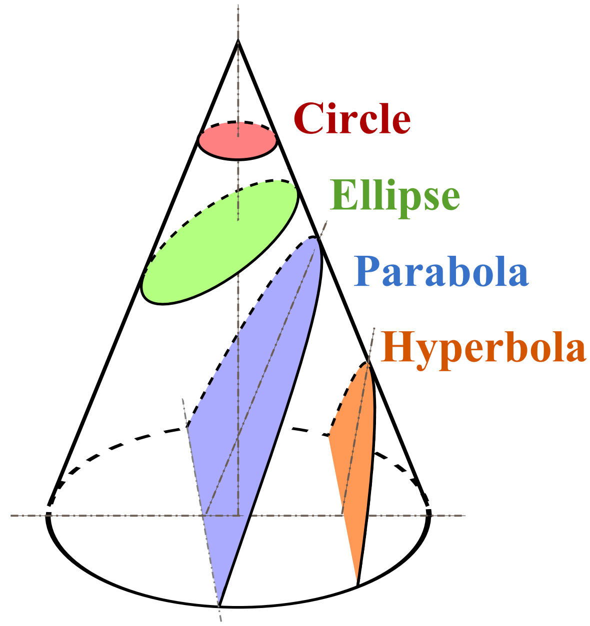
Exercise 11.5.2 *Consider the curve of 11.4.6 in section 11.4. Find the area under one arch of the curve.*

Exercise 11.5.3 *Consider the curve of 11.4.6 in section 11.4. Set up an integral for the length of one arch of the curve.*

11.6 Conics in Polar Coordinates

A **conic section** is a curve obtained as the intersection of a cone and a plane. One useful geometric definition that only involves the plane is that a conic consists of those points whose distances to some point, called a **focus**, and some line, called a **directrix**, are in a fixed ratio, called the **eccentricity**.

The three types of conic sections are the ellipse, parabola and hyperbola (with the circle being a degenerate case of an ellipse).



Let F be a fixed point (the focus), L be a line (the directrix) not containing F and e be a nonnegative real number (the eccentricity). The conics sections are obtained by the set of all points P whose distance to F equals e times their distance to L, that is:

$$\frac{|PF|}{|PL|} = e.$$

In the case that:

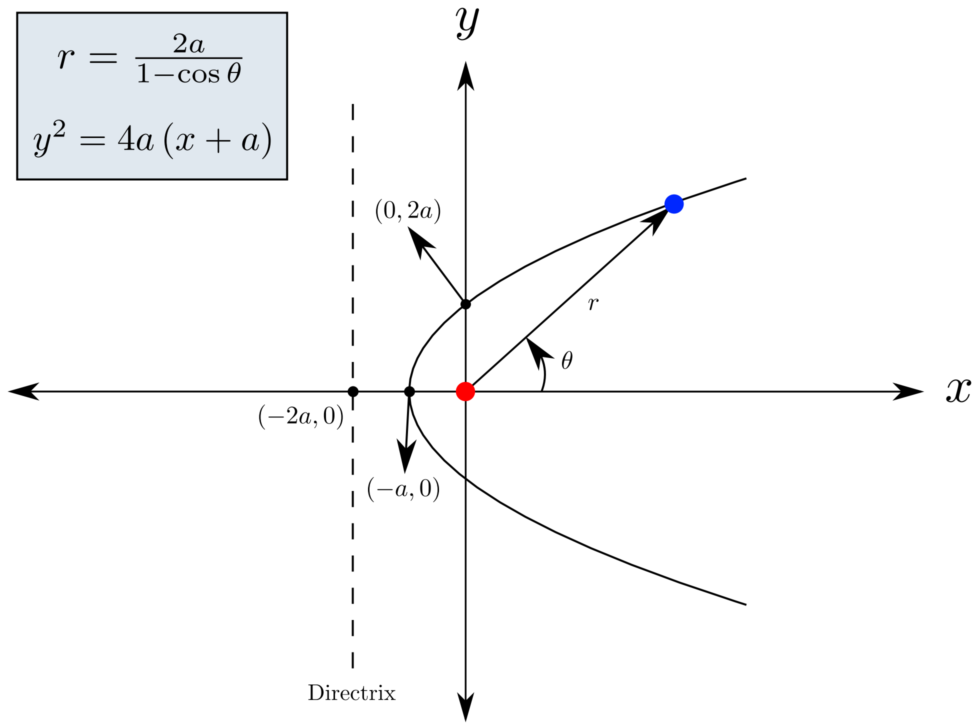
- $e < 1$ we obtain an ellipse (and when $e = 0$ we obtain the degenerate case: A circle),
- $e = 1$ we obtain a parabola,
- $e > 1$ we obtain a hyperbola.

To obtain a simple polar equation we place the focal point at the origin. The formulation for a conic section is then given in the polar form by

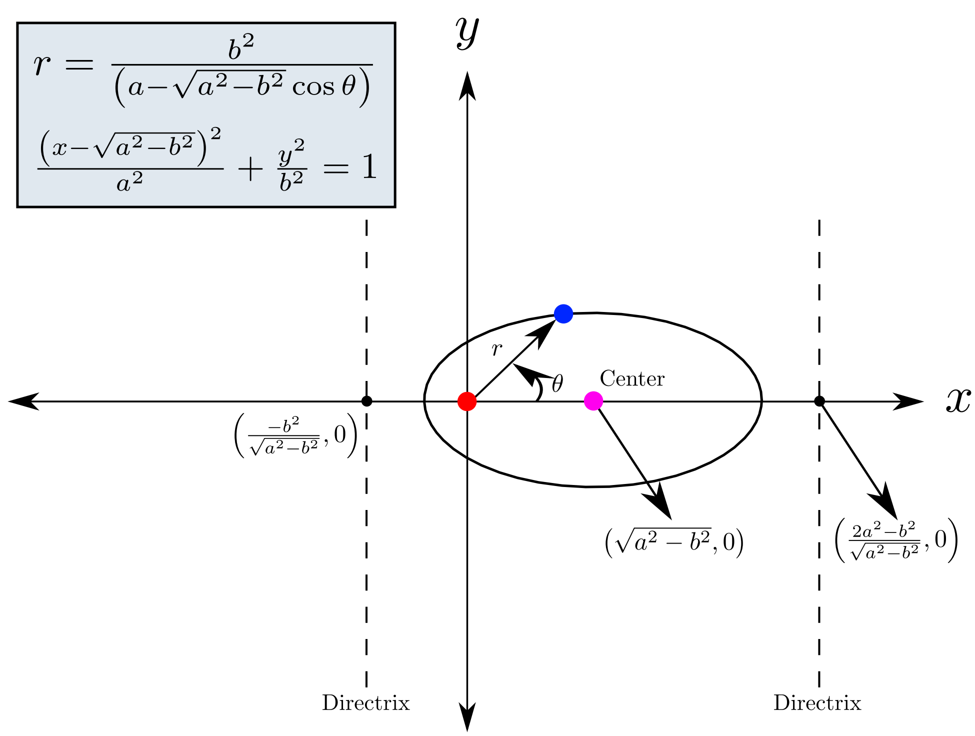
$$r = \frac{pe}{1 \pm e\cos\theta} \qquad \text{and} \qquad r = \frac{pe}{1 \pm e\sin\theta}$$

where e is the eccentricity and p is the focal parameter representing the distance from the focus (or one of the two foci) to the directrix.

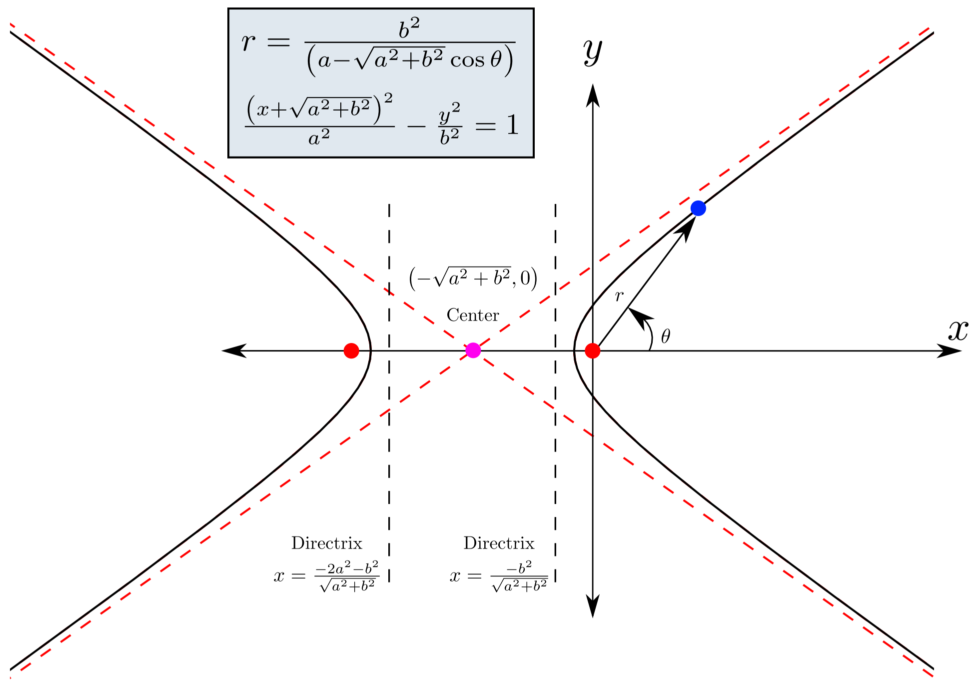
The three different types of conic sections are shown below. Focal points corresponding to all conic sections are placed at the origin. First is the parabola.



Next is the ellipse.



Finally, we have the hyperbola.



Some things to keep in mind with respect to the denominator of polar equations of various conics are the following:

- If the denominator is $1 + e\sin\theta$, it has a horizontal directrix above the focal point.
- If the denominator is $1 - e\sin\theta$, it has a horizontal directrix below the focal point.
- If the denominator is $1 + e\cos\theta$, it has a vertical directrix to the right of the focal point.
- If the denominator is $1 - e\cos\theta$, it has a vertical directrix to the left of the focal point.

Example 11.18: Polar Equations for a Parabola

Find the equation of a parabola with focus at the origin and whose directrix is the line $x = -1$.

Solution. Since we have a parabola, $e = 1$. Furthermore, $p = 1$. Since the graph has a vertical directrix, the equation will use $1 - e\cos\theta$ in the denominator. Thus, the equation is:

$$r = \frac{2}{1 - \sin\theta}$$

♣

Exercises for 11.6

Exercise 11.6.1 *Identify the following conics and find the eccentricity.*

(a) $r = \dfrac{2}{1+\sin\theta}$

(b) $r = \dfrac{4}{2+\cos\theta}$

(c) $r = \dfrac{3}{1-\sin\theta}$

(d) $r = \dfrac{5}{2+2\sin\theta}$

Exercise 11.6.2 *Write the polar equation of a parabola with focus at the origin and directrix* $x = 3$.

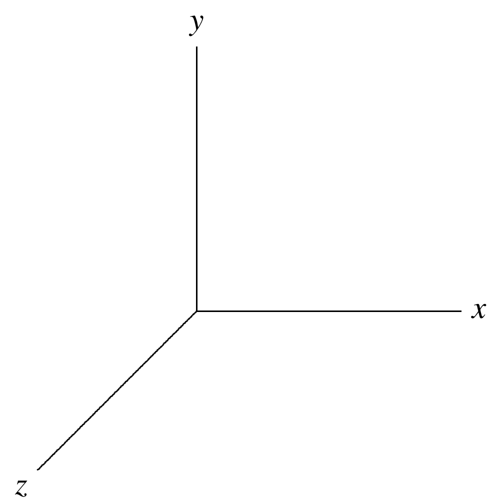
Exercise 11.6.3 *Write the polar equation of a hyperbola with focus at the origin, directrix* $x = 4$ *and eccentricity* 2.

Exercise 11.6.4 *Write the polar equation of an ellipse with focus at the origin, directrix* $x = 4\sec\theta$ *and eccentricity* $1/2$.

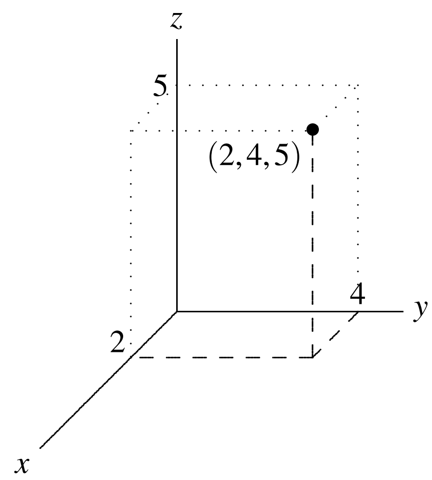
12. Three Dimensions

12.1 The Coordinate System

Throughout the text thus far we have focused investigating functions of the form $y = f(x)$, with one independent and one dependent variable. Such functions can be represented in two dimensions, using two numerical axes that allow us to identify every point in the plane with two numbers. We now shift our focus to three-dimensional space; to identify every point in three dimensions we require three numerical values. The obvious way to make this association is to add one new axis, perpendicular to the x and y axes we already understand. We could, for example, add a third axis, the z axis, with the positive z axis coming straight out of the page, and the negative z axis going out the back of the page. This is difficult to work with on a printed page, so more often we draw a view of the three axes from an angle:



You must then imagine that the z axis is perpendicular to the other two. Just as we have investigated functions of the form $y = f(x)$ in two dimensions, we will investigate three dimensions largely by considering functions; now the functions will (typically) have the form $z = f(x, y)$. Due to the fact that we are used to having the result of a function graphed in the vertical direction, it is somewhat easier to maintain that convention in three dimensions. To accomplish this, we normally rotate the axes so that z points up; the result is then:



Note that if you imagine looking down from above, along the z axis, the positive z axis will come straight toward you, the positive y axis will point up, and the positive x axis will point to your right, as usual. Any point in space is identified by providing the three coordinates of the point, as shown; naturally, we list the coordinates in the order (x, y, z). One useful way to think of this is to use the x and y coordinates to identify a point in the x-y plane, then move straight up (or down) a distance given by the z coordinate.

It is now fairly simple to understand some "shapes" in three dimensions that correspond to simple conditions on the coordinates. In two dimensions the equation $x = 1$ describes the vertical line through $(1, 0)$. In three dimensions, it still describes all points with x-coordinate 1, but this is now a plane, as in Figure 12.1.

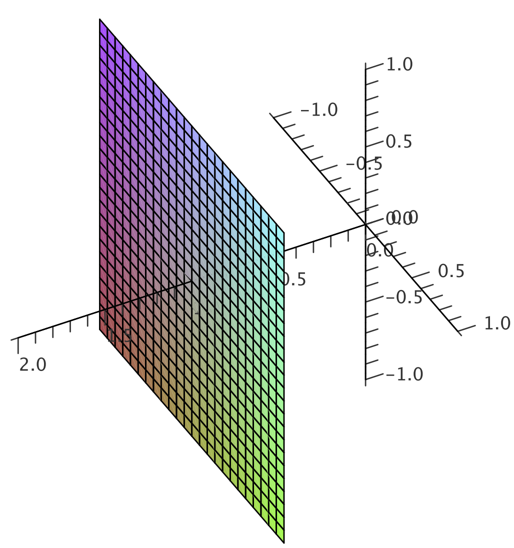


Figure 12.1: The plane $x = 1$.

Recall the very useful distance formula in two dimensions which comes directly from the Pythagorean Theorem: the distance between points (x_1, y_1) and (x_2, y_2) is $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. What is the distance between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) in three dimensions? Geometrically, we want the length of the long diagonal labelled c in the "box" in Figure 12.2. Since a, b, c form a right triangle, $a^2 + b^2 = c^2$. b is the vertical distance between (x_1, y_1, z_1) and (x_2, y_2, z_2), so $b = |z_1 - z_2|$. The length a runs parallel to the x-y plane, so it is simply the distance between (x_1, y_1) and (x_2, y_2), that is, $a^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$. Now we see that $c^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$ and $c = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$.

It is sometimes useful to give names to points, for example we might let $P_1 = (x_1, y_1, z_1)$, or more concisely we might refer to the point $P_1(x_1, y_1, z_1)$, and subsequently use just P_1. Distance between two points in either two or three dimensions is sometimes denoted by d, so for example the formula for the distance between $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ might be expressed as

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

Distance

The distance between points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ in two dimensions is

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

The distance between points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ in three dimensions is

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

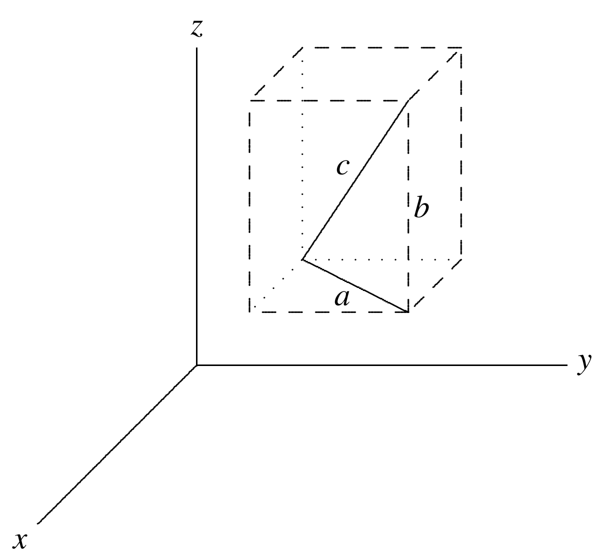


Figure 12.2: Distance in three dimensions.

In two dimensions, the distance formula immediately gives us the equation of a circle: the circle of radius r and center at (h, k) consists of all points (x, y) at distance r from (h, k), so the equation is $r = \sqrt{(x-h)^2 + (y-k)^2}$ or $r^2 = (x-h)^2 + (y-k)^2$. Now we can get the similar equation $r^2 = (x-h)^2 + (y-k)^2 + (z-l)^2$, which describes all points (x, y, z) at distance r from (h, k, l), namely, the sphere with radius r and center (h, k, l).

Exercises for 12.1

Exercise 12.1.1 *Sketch the location of the points* $(1,1,0)$, $(2,3,-1)$, *and* $(-1,2,3)$ *on a single set of axes.*

Exercise 12.1.2 *Describe geometrically the set of points* (x,y,z) *that satisfy* $z=4$.

Exercise 12.1.3 *Describe geometrically the set of points* (x,y,z) *that satisfy* $y=-3$.

Exercise 12.1.4 *Describe geometrically the set of points* (x,y,z) *that satisfy* $x+y=2$.

Exercise 12.1.5 *The equation* $x+y+z=1$ *describes some collection of points in* $\mathbb{R}^3$. *Describe and sketch the points that satisfy* $x+y+z=1$ *and are in the x-y plane, in the x-z plane, and in the y-z plane.*

Exercise 12.1.6 *Find the lengths of the sides of the triangle with vertices* $(1,0,1)$, $(2,2,-1)$, *and* $(-3,2,-2)$.

Exercise 12.1.7 *Find the lengths of the sides of the triangle with vertices* $(2,2,3)$, $(8,6,5)$, *and* $(-1,0,2)$. *Why do the results tell you that this isn't really a triangle?*

Exercise 12.1.8 *Find an equation of the sphere with center at* $(1,1,1)$ *and radius 2.*

Exercise 12.1.9 *Find an equation of the sphere with center at* $(2,-1,3)$ *and radius 5.*

Exercise 12.1.10 *Find an equation of the sphere with center* $(3,-2,1)$ *and that goes through the point* $(4,2,5)$.

Exercise 12.1.11 *Find an equation of the sphere with center at* $(2,1,-1)$ *and radius 4. Find an equation for the intersection of this sphere with the y-z plane; describe this intersection geometrically.*

Exercise 12.1.12 *Consider the sphere of radius 5 centered at* $(2,3,4)$. *What is the intersection of this sphere with each of the coordinate planes?*

Exercise 12.1.13 *Show that for all values of* θ *and* ϕ, *the point* $(a\sin\phi\cos\theta, a\sin\phi\sin\theta, a\cos\phi)$ *lies on the sphere given by* $x^2+y^2+z^2=a^2$.

Exercise 12.1.14 *Prove that the midpoint of the line segment connecting* (x_1,y_1,z_1) *to* (x_2,y_2,z_2) *is at* $\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}, \frac{z_1+z_2}{2}\right)$.

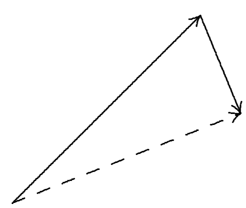
Exercise 12.1.15 *Any three points* $P_1(x_1,y_1,z_1)$, $P_2(x_2,y_2,z_2)$, $P_3(x_3,y_3,z_3)$, *lie in a plane and form a triangle. The* **triangle inequality** *says that* $d(P_1,P_3)\le d(P_1,P_2)+d(P_2,P_3)$. *Prove the triangle inequality using either algebra (messy) or the law of cosines (less messy).*

Exercise 12.1.16 *Is it possible for a plane to intersect a sphere in exactly two points? Exactly one point? Explain.*

12.2 Vectors

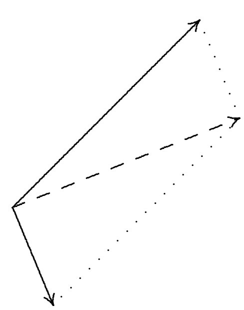
A **vector**, denoted **v**, is a quantity consisting of a non-negative magnitude and a direction. We could represent a vector in two dimensions as (m, θ), where m is the magnitude and θ is the direction, measured as an angle from some agreed upon direction. For example, we might think of the vector $(5, 45°)$ as representing "5 km toward the northeast"; that is, this vector might be a **displacement vector**, indicating, say, that your grandfather walked 5 kilometers toward the northeast to school in the snow. On the other hand, the same vector could represent a velocity, indicating that your grandfather walked at 5 km/hr toward the northeast. What the vector does not indicate is where this walk occurred: a vector represents a magnitude and a direction, but not a location. Pictorially it is useful to represent a vector as an arrow; the direction of the vector, naturally, is the direction in which the arrow points; the magnitude of the vector is reflected in the length of the arrow.

It turns out that many, many quantities behave as vectors, e.g., displacement, velocity, acceleration, force. Already we can get some idea of their usefulness using displacement vectors. Suppose that your grandfather walked 5 km NE and then 2 km SSE; if the terrain allows, and perhaps armed with a compass, how could your grandfather have walked directly to his destination? We can use vectors (and a bit of geometry) to answer this question. We begin by noting that since vectors do not include a specification of position, we can "place" them anywhere that is convenient. So we can picture your grandfather's journey as two displacement vectors drawn head to tail:



The displacement vector for the shortcut route is the vector drawn with a dashed line, from the tail of the first to the head of the second. With a little trigonometry, we can compute that the third vector has magnitude approximately 4.62 and direction 21.43°, so walking 4.62 km in the direction 21.43° north of east (approximately ENE) would get your grandfather to school. This sort of calculation is so common, we dignify it with a name: we say that the third vector is the **sum** of the other two vectors.

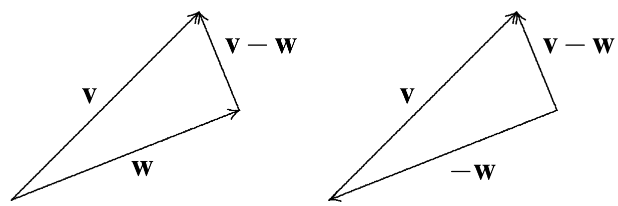
There is another common way to picture the sum of two vectors. Put the vectors tail to tail and then complete the parallelogram they indicate; the sum of the two vectors is the diagonal of the parallelogram:



This is a more natural representation in some circumstances. For example, if the two original vectors represent forces acting on an object, the sum of the two vectors is the net or effective force on the object, and it is convenient to draw all three with their tails at the location of the object.

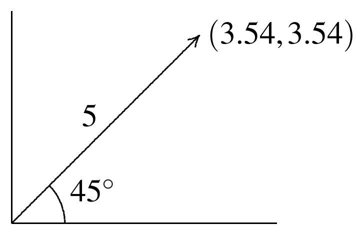
We also define **scalar multiplication** for vectors: if $\mathbf{v}$ is a vector (m, θ) and $a \geq 0$ is a real number, the vector $a\mathbf{v}$ is (am, θ), namely, it points in the same direction but has a times the magnitude. If $a < 0$, $a\mathbf{v}$ is $(|a|m, \theta + \pi)$, with $|a|$ times the magnitude and pointing in the opposite direction (unless we specify otherwise, angles are measured in radians).

Now we can understand subtraction of vectors: $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-1)\mathbf{w}$:

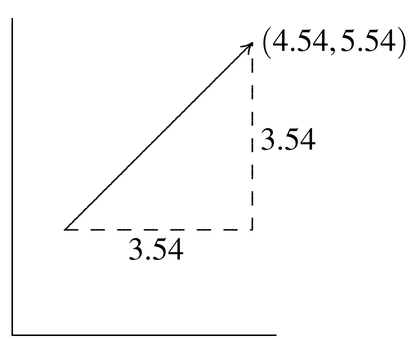


Note that as you would expect, $\mathbf{w} + (\mathbf{v} - \mathbf{w}) = \mathbf{v}$.

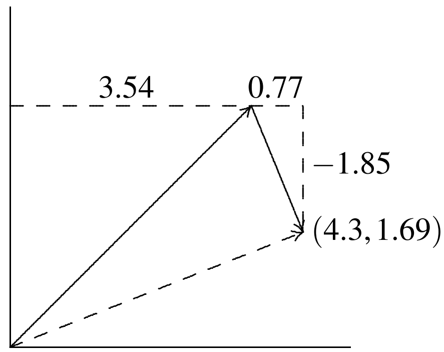
We can represent a vector in ways other than (m, θ), and in fact (m, θ) is not generally used at all. How else could we describe a particular vector? Consider again the vector $(5, 45^\circ)$. Let's draw it again, but impose a coordinate system. If we put the tail of the arrow at the origin, the head of the arrow ends up at the point $(5/\sqrt{2}, 5/\sqrt{2}) \approx (3.54, 3.54)$.



In this picture the coordinates $(3.54, 3.54)$ identify the head of the arrow, provided we know that the tail of the arrow has been placed at $(0,0)$. Then in fact the vector can always be identified as $(3.54, 3.54)$, no matter where it is placed; we just have to remember that the numbers 3.54 must be interpreted as a *change* from the position of the tail, not as the actual coordinates of the arrow head; to emphasize this we will write $\langle 3.54, 3.54 \rangle$ to mean the vector and $(3.54, 3.54)$ to mean the point. Then if the vector $\langle 3.54, 3.54 \rangle$ is drawn with its tail at $(1,2)$ it looks like this:



Consider again the two part trip: 5 km NE and then 2 km SSE. The vector representing the first part of the trip is $\langle 5/\sqrt{2}, 5/\sqrt{2}\rangle$, and the second part of the trip is represented by $\langle 2\cos(-3\pi/8), 2\sin(-3\pi/8)\rangle \approx \langle 0.77, -1.85\rangle$. We can represent the sum of these with the usual head to tail picture:



It is clear from the picture that the coordinates of the destination point are $(5/\sqrt{2}+2\cos(-3\pi/8), 5/\sqrt{2}+2\sin(-3\pi/8))$ or approximately $(4.3, 1.69)$, so the sum of the two vectors is $\langle 5/\sqrt{2}+2\cos(-3\pi/8), 5/\sqrt{2}+2\sin(-3\pi/8)\rangle \approx \langle 4.3, 1.69\rangle$. Adding the two vectors is easier in this form than in the (m,θ) form, provided that we're willing to have the answer in this form as well.

It is easy to see that scalar multiplication and vector subtraction are also easy to compute in this form: $a\langle v,w\rangle = \langle av, aw\rangle$ and $\langle v_1,w_1\rangle - \langle v_2,w_2\rangle = \langle v_1-v_2, w_1-w_2\rangle$. What about the magnitude? The magnitude of the vector $\langle v,w\rangle$ is still the length of the corresponding arrow representation; this is the distance from the origin to the point (v,w), namely, the distance from the tail to the head of the arrow. Using the familiar distance formula the magnitude of the vector is simply $\sqrt{v^2+w^2}$, which we also denote with absolute value bars: $|\langle v,w\rangle| = \sqrt{v^2+w^2}$.

In three dimensions, vectors are still quantities consisting of a magnitude and a direction, but of course there are many more possible directions. It's not clear how we might represent the direction explicitly, but the coordinate version of vectors makes just as much sense in three dimensions as in two. By $\langle 1,2,3\rangle$ we mean the vector whose head is at $(1,2,3)$ if its tail is at the origin. As before, we can place the vector anywhere we want; if it has its tail at $(4,5,6)$ then its head is at $(5,7,9)$. It remains true that arithmetic is easy to do with vectors in this form.

Arithmetic of Vectors

Sum of vectors:

$$\langle v_1,v_2,v_3\rangle + \langle w_1,w_2,w_3\rangle = \langle v_1+w_1, v_2+w_2, v_3+w_3\rangle$$

Scalar Multiplication of vectors:

$$a\langle v_1,v_2,v_3\rangle = \langle av_1, av_2, av_3\rangle$$

Subtraction of vectors:

$$\langle v_1,v_2,v_3\rangle - \langle w_1,w_2,w_3\rangle = \langle v_1-w_1, v_2-w_2, v_3-w_3\rangle$$

The **magnitude** of the vector is the distance from the origin to the head of the vector, or $|\langle v_1,v_2,v_3\rangle| = \sqrt{v_1^2+v_2^2+v_3^2}$.

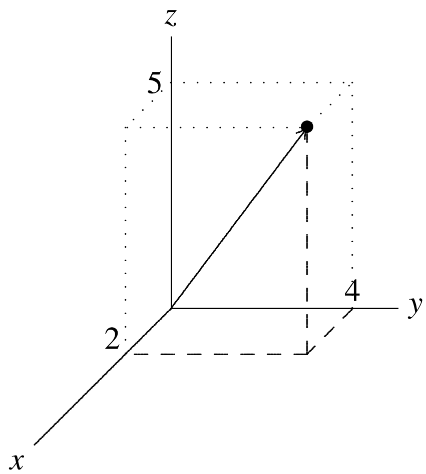


Figure 12.3: The vector $\langle 2,4,5\rangle$ with its tail at the origin.

Three particularly simple vectors turn out to be quite useful: $\mathbf{i}=\langle 1,0,0\rangle$, $\mathbf{j}=\langle 0,1,0\rangle$, and $\mathbf{k}=\langle 0,0,1\rangle$. These play much the same role for vectors that the axes play for points. In particular, notice that

$$\begin{aligned}\langle v_1,v_2,v_3\rangle &= \langle v_1,0,0\rangle+\langle 0,v_2,0\rangle+\langle 0,0,v_3\rangle\\ &= v_1\langle 1,0,0\rangle+v_2\langle 0,1,0\rangle+v_3\langle 0,0,1\rangle\\ &= v_1\mathbf{i}+v_2\mathbf{j}+v_3\mathbf{k}\end{aligned}$$

Thus far, we have focused our discussion on vectors which begin at the origin and end at a point. However we will frequently want to produce a vector that points from one point to another. That is, if P and Q are points, we seek the vector $\mathbf{x}$ such that when the tail of $\mathbf{x}$ is placed at P, its head is at Q; we refer to this vector as $\overrightarrow{PQ}$. If we know the coordinates of P and Q, the coordinates of the vector are easy to find.

Example 12.1:

Suppose $P=(1,-2,4)$ *and* $Q=(-2,1,3)$. *The vector* $\overrightarrow{PQ}$ *is* $\langle -2-1, 1--2, 3-4\rangle=\langle -3,3,-1\rangle$ *and* $\overrightarrow{QP}=\langle 3,-3,1\rangle$.

Exercises for 12.2

Exercise 12.2.1 *Draw the vector* $\langle 3,-1\rangle$ *with its tail at the origin.*

Exercise 12.2.2 *Draw the vector* $\langle 3,-1,2\rangle$ *with its tail at the origin.*

Exercise 12.2.3 *Let* $\mathbf{v}$ *be the vector with tail at the origin and head at* $(1,2)$*; let* $\mathbf{w}$ *be the vector with tail at the origin and head at* $(3,1)$. *Draw* $\mathbf{v}$ *and* $\mathbf{w}$ *and a vector* $\mathbf{u}$ *with tail at* $(1,2)$ *and head at* $(3,1)$. *Draw* $\mathbf{u}$ *with its tail at the origin.*

Exercise 12.2.4 *Let* $\mathbf{v}$ *be the vector with tail at the origin and head at* $(-1,2)$*; let* $\mathbf{w}$ *be the vector with tail at the origin and head at* $(3,3)$*. Draw* $\mathbf{v}$ *and* $\mathbf{w}$ *and a vector* $\mathbf{u}$ *with tail at* $(-1,2)$ *and head at* $(3,3)$*. Draw* $\mathbf{u}$ *with its tail at the origin.*

Exercise 12.2.5 *Let* $\mathbf{v}$ *be the vector with tail at the origin and head at* $(5,2)$*; let* $\mathbf{w}$ *be the vector with tail at the origin and head at* $(1,5)$*. Draw* $\mathbf{v}$ *and* $\mathbf{w}$ *and a vector* $\mathbf{u}$ *with tail at* $(5,2)$ *and head at* $(1,5)$*. Draw* $\mathbf{u}$ *with its tail at the origin.*

Exercise 12.2.6 *Find* $|\mathbf{v}|$, $\mathbf{v}+\mathbf{w}$, $\mathbf{v}-\mathbf{w}$, $|\mathbf{v}+\mathbf{w}|$, $|\mathbf{v}-\mathbf{w}|$ *and* $-2\mathbf{v}$ *for* $\mathbf{v}=\langle 1,3\rangle$ *and* $\mathbf{w}=\langle -1,-5\rangle$*.*

Exercise 12.2.7 *Find* $|\mathbf{v}|$, $\mathbf{v}+\mathbf{w}$, $\mathbf{v}-\mathbf{w}$, $|\mathbf{v}+\mathbf{w}|$, $|\mathbf{v}-\mathbf{w}|$ *and* $-2\mathbf{v}$ *for* $\mathbf{v}=\langle 1,2,3\rangle$ *and* $\mathbf{w}=\langle -1,2,-3\rangle$*.*

Exercise 12.2.8 *Find* $|\mathbf{v}|$, $\mathbf{v}+\mathbf{w}$, $\mathbf{v}-\mathbf{w}$, $|\mathbf{v}+\mathbf{w}|$, $|\mathbf{v}-\mathbf{w}|$ *and* $-2\mathbf{v}$ *for* $\mathbf{v}=\langle 1,0,1\rangle$ *and* $\mathbf{w}=\langle -1,-2,2\rangle$*.*

Exercise 12.2.9 *Find* $|\mathbf{v}|$, $\mathbf{v}+\mathbf{w}$, $\mathbf{v}-\mathbf{w}$, $|\mathbf{v}+\mathbf{w}|$, $|\mathbf{v}-\mathbf{w}|$ *and* $-2\mathbf{v}$ *for* $\mathbf{v}=\langle 1,-1,1\rangle$ *and* $\mathbf{w}=\langle 0,0,3\rangle$*.*

Exercise 12.2.10 *Find* $|\mathbf{v}|$, $\mathbf{v}+\mathbf{w}$, $\mathbf{v}-\mathbf{w}$, $|\mathbf{v}+\mathbf{w}|$, $|\mathbf{v}-\mathbf{w}|$ *and* $-2\mathbf{v}$ *for* $\mathbf{v}=\langle 3,2,1\rangle$ *and* $\mathbf{w}=\langle -1,-1,-1\rangle$*.*

Exercise 12.2.11 *Let* $P=(4,5,6)$, $Q=(1,2,-5)$*. Find* $\overrightarrow{PQ}$*. Find a vector with the same direction as* $\overrightarrow{PQ}$ *but with length 1. Find a vector with the same direction as* $\overrightarrow{PQ}$ *but with length 4.*

Exercise 12.2.12 *If* $A,B,$ *and* C *are three points, find* $\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CA}$*.*

Exercise 12.2.13 *Consider the 12 vectors that have their tails at the center of a clock and their respective heads at each of the 12 digits. What is the sum of these vectors? What if we remove the vector corresponding to 4 o'clock? What if, instead, all vectors have their tails at 12 o'clock, and their heads on the remaining digits?*

Exercise 12.2.14 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be nonzero vectors in two dimensions that are not parallel or anti-parallel. Show, algebraically, that if* $\mathbf{c}$ *is any two dimensional vector, there are scalars* s *and* t *such that* $\mathbf{c}=s\mathbf{a}+t\mathbf{b}$*.*

Exercise 12.2.15 *Does the statement in the previous exercise hold if the vectors* $\mathbf{a}$, $\mathbf{b}$, *and* $\mathbf{c}$ *are three dimensional vectors? Explain.*

12.3 The Dot Product

The goal of this section is to answer the following question. Given two vectors, what is the angle between them?

Since vectors have no position, we are free to place vectors wherever we like. If the two vectors are placed tail-to-tail, there is now a reasonable interpretation of the question: we seek the measure of the smallest angle between the two vectors, in the plane in which they lie. Figure 12.4 illustrates the situation.

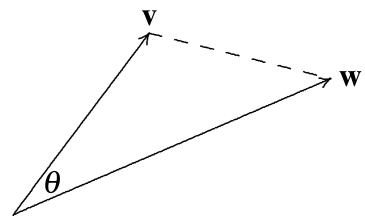


Figure 12.4: The angle between vectors v and w.

Since the angle θ lies in a triangle, we can compute it using a bit of trigonometry, namely, the law of cosines. Remember that the law of cosines states $c^2 = a^2 + b^2 - 2ab\cos C$.

The lengths of the sides of the triangle in Figure 12.4 are $|\mathbf{v}|$, $|\mathbf{w}|$, and $|\mathbf{v}-\mathbf{w}|$. Let $\mathbf{v} = \langle v_1, v_2, v_3\rangle$ and $\mathbf{w} = \langle w_1, w_2, w_3\rangle$; then

$$\begin{aligned}
|\mathbf{v}-\mathbf{w}|^2 &= |\mathbf{v}|^2 + |\mathbf{w}|^2 - 2|\mathbf{v}||\mathbf{w}|\cos\theta\\
2|\mathbf{v}||\mathbf{w}|\cos\theta &= |\mathbf{v}|^2 + |\mathbf{w}|^2 - |\mathbf{v}-\mathbf{w}|^2\\
&= v_1^2+v_2^2+v_3^2+w_1^2+w_2^2+w_3^2-(v_1-w_1)^2-(v_2-w_2)^2-(v_3-w_3)^2\\
&= v_1^2+v_2^2+v_3^2+w_1^2+w_2^2+w_3^2\\
&\qquad -(v_1^2-2v_1w_1+w_1^2)-(v_2^2-2v_2w_2+w_2^2)-(v_3^2-2v_3w_3+w_3^2)\\
&= 2v_1w_1+2v_2w_2+2v_3w_3\\
|\mathbf{v}||\mathbf{w}|\cos\theta &= v_1w_1+v_2w_2+v_3w_3\\
\cos\theta &= (v_1w_1+v_2w_2+v_3w_3)/(|\mathbf{v}||\mathbf{w}|)
\end{aligned}$$

A bit of simple arithmetic with the coordinates of $\mathbf{v}$ and $\mathbf{w}$ allows us to compute the cosine of the angle between them. If necessary we can use the arccosine to get θ, but in many problems $\cos\theta$ turns out to be all we really need.

The numerator of the fraction that gives us $\cos\theta$ turns up a lot, so we give it a name and more compact notation: we call it the **dot product**, and write it as

$$\mathbf{v}\cdot\mathbf{w} = v_1w_1+v_2w_2+v_3w_3$$

This is the same symbol we use for ordinary multiplication, but there should never be any confusion; you can tell from context whether we are "multiplying" vectors or numbers. (We might also use the dot for scalar multiplication: $a\cdot\mathbf{v} = a\mathbf{v}$; again, it is clear what is meant from context.)

Example 12.2:

Find the angle between the vectors $\mathbf{v} = \langle 1,2,1\rangle$ *and* $\mathbf{w} = \langle 3,1,-5\rangle$.

Solution. We know that $\cos\theta = \mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|) = (1\cdot 3+2\cdot 1+1\cdot(-5))/(|\mathbf{v}||\mathbf{w}|) = 0$, so $\theta = \pi/2$, that is, the vectors are perpendicular. ♣

Example 12.3:

Find the angle between the vectors $\mathbf{v} = \langle 3,3,0\rangle$ *and* $\mathbf{w} = \langle 1,0,0\rangle$.

Solution. We compute

$$\begin{aligned}\cos\theta &= (3\cdot 1+3\cdot 0+0\cdot 0)/(\sqrt{9+9+0}\sqrt{1+0+0})\\ &= 3/\sqrt{18} = 1/\sqrt{2}\end{aligned}$$

so $\theta = \pi/4$. ♣

The following are some special cases worth looking at.

Example 12.4:

Find the angles between:

1. $\mathbf{v}$ *and* $\mathbf{v}$
2. $\mathbf{v}$ *and* $-\mathbf{v}$
3. $\mathbf{v}$ *and* $\mathbf{0} = \langle 0,0,0\rangle$

Solution.

1. $\cos\theta = \mathbf{v}\cdot\mathbf{v}/(|\mathbf{v}||\mathbf{v}|) = (v_1^2+v_2^2+v_3^2)/(\sqrt{v_1^2+v_2^2+v_3^2}\sqrt{v_1^2+v_2^2+v_3^2}) = 1$, so the angle between $\mathbf{v}$ and itself is zero, which of course is correct.
2. $\cos\theta = \mathbf{v}\cdot -\mathbf{v}/(|\mathbf{v}||-\mathbf{v}|) = (-v_1^2-v_2^2-v_3^2)/(\sqrt{v_1^2+v_2^2+v_3^2}\sqrt{v_1^2+v_2^2+v_3^2}) = -1$, so the angle is π, that is, the vectors point in opposite directions, as of course we already knew.
3. $\cos\theta = \mathbf{v}\cdot\mathbf{0}/(|\mathbf{v}||\mathbf{0}|) = (0+0+0)/(\sqrt{v_1^2+v_2^2+v_3^2}\sqrt{0^2+0^2+0^2})$, which is undefined. On the other hand, note that since $\mathbf{v}\cdot\mathbf{0} = 0$ it looks at first as if $\cos\theta$ will be zero, which as we have seen means that vectors are perpendicular; only when we notice that the denominator is also zero do we run into trouble. One way to "fix" this is to adopt the convention that the zero vector $\mathbf{0}$ is perpendicular to all vectors; then we can say in general that if $\mathbf{v}\cdot\mathbf{w} = 0$, $\mathbf{v}$ and $\mathbf{w}$ are perpendicular.

♣

Generalizing the examples, note the following useful facts:

- If $\mathbf{v}$ is parallel or anti-parallel to $\mathbf{w}$ then $\mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|) = \pm 1$, and conversely, if $\mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|) = 1$, $\mathbf{v}$ and $\mathbf{w}$ are parallel, while if $\mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|) = -1$, $\mathbf{v}$ and $\mathbf{w}$ are anti-parallel. (Vectors are parallel if they point in the same direction, anti-parallel if they point in opposite directions.)
- If $\mathbf{v}$ is perpendicular to $\mathbf{w}$ then $\mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|) = 0$, and conversely if $\mathbf{v}\cdot\mathbf{w}/(|\mathbf{v}||\mathbf{w}|) = 0$ then $\mathbf{v}$ and $\mathbf{w}$ are perpendicular.

Given two vectors, it is often useful to find the **projection** of one vector onto the other, because this turns out to have important meaning in many circumstances. More precisely, given $\mathbf{v}$ and $\mathbf{w}$, we seek a vector parallel to $\mathbf{w}$ but with length determined by $\mathbf{v}$ in a natural way, as shown in Figure 12.5. $\mathbf{p}$ is chosen so that the triangle formed by $\mathbf{v}$, $\mathbf{p}$, and $\mathbf{v}-\mathbf{p}$ is a right triangle.

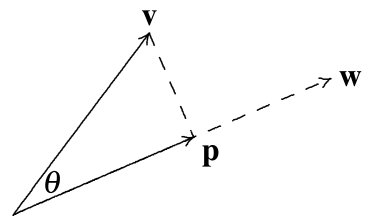


Figure 12.5: p is the projection of v onto w.

Using a little trigonometry, we see that

$$|\mathbf{p}| = |\mathbf{v}|\cos\theta = |\mathbf{v}|\frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{v}||\mathbf{w}|} = \frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{w}|};$$

this is sometimes called the **scalar projection of v onto w**. To get **p** itself, we multiply this length by a vector of length one parallel to **w**:

$$\mathbf{p} = \frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{w}|}\frac{\mathbf{w}}{|\mathbf{w}|} = \frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{w}|^2}\mathbf{w}.$$

Be sure that you understand why $\mathbf{w}/|\mathbf{w}|$ is a vector of length one (also called a **unit vector**) parallel to **w**.

The discussion so far implicitly assumed that $0 \le \theta \le \pi/2$. If $\pi/2 < \theta \le \pi$, the picture is like Figure 12.6. In this case $\mathbf{v}\cdot\mathbf{w}$ is negative, so the vector

$$\frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{w}|^2}\mathbf{w}$$

is anti-parallel to **w**, and its length is

$$\left|\frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{w}|}\right|.$$

In general, the scalar projection of **v** onto **w** may be positive or negative. If it is negative, it means that the projection vector is anti-parallel to **w** and that the length of the projection vector is the absolute value of the scalar projection. Of course, you can also compute the length of the projection vector as usual, by applying the distance formula to the vector.

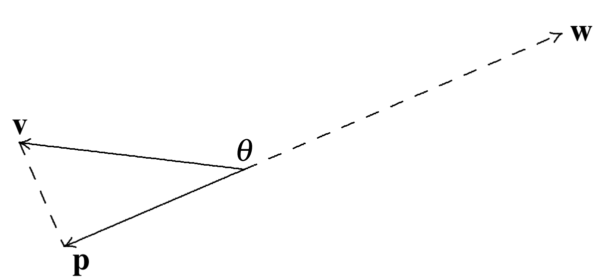


Figure 12.6: p is the projection of v onto w.

Note that the phrase "projection onto **w**" is a bit misleading if taken literally; all that **w** provides is a direction; the length of **w** has no impact on the final vector. In Figure 12.7, for example, **w** is shorter than the projection vector, but this is perfectly acceptable.

Figure 12.7: p is the projection of v onto w.

Physical force is a vector quantity. It is often necessary to compute the "component" of a force acting in a different direction than the force is being applied.

Example 12.5: Components of Force Vector

Suppose a ten pound weight is resting on an inclined plane—a pitched roof, for example. Gravity exerts a force of ten pounds on the object, directed straight down. It is useful to think of the component of this force directed down and parallel to the roof, and the component down and directly into the roof. These forces are the projections of the force vector onto vectors parallel and perpendicular to the roof. Suppose the roof is tilted at a 30° *angle, as in Figure 12.8. Compute the component of the force directed down the roof and the component of the force directed into the roof.*

Solution. A vector parallel to the roof is $\langle -\sqrt{3}, -1\rangle$, and a vector perpendicular to the roof is $\langle 1, -\sqrt{3}\rangle$. The force vector is $\mathbf{F} = \langle 0, -10\rangle$. The component of the force directed down the roof is then

$$\mathbf{F}_1 = \frac{\mathbf{F}\cdot\langle -\sqrt{3}, -1\rangle}{|\langle -\sqrt{3}, -1\rangle|^2}\langle -\sqrt{3}, -1\rangle = \frac{10}{2}\frac{\langle -\sqrt{3}, -1\rangle}{2} = \langle -5\sqrt{3}/2, -5/2\rangle$$

with length 5. The component of the force directed into the roof is

$$\mathbf{F}_2 = \frac{\mathbf{F}\cdot\langle 1, -\sqrt{3}\rangle}{|\langle 1, -\sqrt{3}\rangle|^2}\langle 1, -\sqrt{3}\rangle = \frac{10\sqrt{3}}{2}\frac{\langle 1, -\sqrt{3}\rangle}{2} = \langle 5\sqrt{3}/2, -15/2\rangle$$

with length $5\sqrt{3}$. Thus, a force of 5 pounds is pulling the object down the roof, while a force of $5\sqrt{3}$ pounds is pulling the object into the roof. ♣

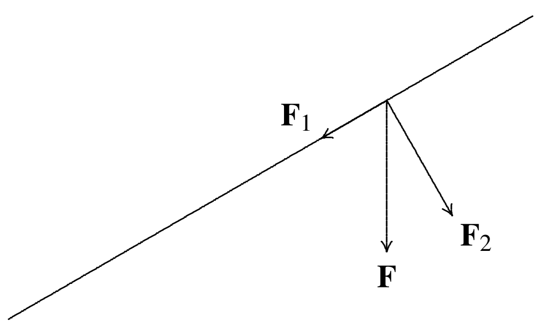


Figure 12.8: Components of a force.

The dot product has some familiar-looking properties that will be useful later, so we list them here. These may be proved by writing the vectors in coordinate form and then performing the indicated calculations; subsequently it can be easier to use the properties instead of calculating with coordinates.

Theorem 12.6: Dot Product Properties

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors and a is a real number, then

1. $\mathbf{u}\cdot\mathbf{u} = |\mathbf{u}|^2$
2. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$
3. $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$
4. $(a\mathbf{u})\cdot\mathbf{v} = a(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(a\mathbf{v})$

Exercises for 12.3

Exercise 12.3.1 *Find* $\langle 1,1,1\rangle\cdot\langle 2,-3,4\rangle$.

Exercise 12.3.2 *Find* $\langle 1,2,0\rangle\cdot\langle 0,0,57\rangle$.

Exercise 12.3.3 *Find* $\langle 3,2,1\rangle\cdot\langle 0,1,0\rangle$.

Exercise 12.3.4 *Find* $\langle -1,-2,5\rangle\cdot\langle 1,0,-1\rangle$.

Exercise 12.3.5 *Find* $\langle 3,4,6\rangle\cdot\langle 2,3,4\rangle$.

Exercise 12.3.6 *Find the cosine of the angle between* $\langle 1,2,3\rangle$ *and* $\langle 1,1,1\rangle$*; use a calculator if necessary to find the angle.*

Exercise 12.3.7 *Find the cosine of the angle between* $\langle -1,-2,-3\rangle$ *and* $\langle 5,0,2\rangle$*; use a calculator if necessary to find the angle.*

Exercise 12.3.8 *Find the cosine of the angle between* $\langle 47,100,0\rangle$ *and* $\langle 0,0,5\rangle$*; use a calculator if necessary to find the angle.*

Exercise 12.3.9 *Find the cosine of the angle between* $\langle 1,0,1\rangle$ *and* $\langle 0,1,1\rangle$*; use a calculator if necessary to find the angle.*

Exercise 12.3.10 *Find the cosine of the angle between* $\langle 2,0,0\rangle$ *and* $\langle -1,1,-1\rangle$*; use a calculator if necessary to find the angle.*

Exercise 12.3.11 *Find the angle between the diagonal of a cube and one of the edges adjacent to the diagonal.*

Exercise 12.3.12 *Find the scalar and vector projections of* $\langle 1,2,3\rangle$ *onto* $\langle 1,2,0\rangle$.

Exercise 12.3.13 *Find the scalar and vector projections of* $\langle 1,1,1\rangle$ *onto* $\langle 3,2,1\rangle$.

Exercise 12.3.14 *A force of 10 pounds is applied to a wagon, directed at an angle of* $30°$. *Find the component of this force pulling the wagon straight up, and the component pulling it horizontally along the ground.*

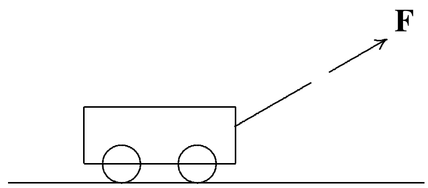


Figure 12.9: Pulling a wagon.

Exercise 12.3.15 *A force of 15 pounds is applied to a wagon, directed at an angle of* $45°$. *Find the component of this force pulling the wagon straight up, and the component pulling it horizontally along the ground.*

Exercise 12.3.16 *Use the dot product to find a non-zero vector* $\boldsymbol{w}$ *perpendicular to both* $\boldsymbol{u} = \langle 1,2,-3\rangle$ *and* $\boldsymbol{v} = \langle 2,0,1\rangle$.

Exercise 12.3.17 *Let* $\boldsymbol{x} = \langle 1,1,0\rangle$ *and* $\boldsymbol{y} = \langle 2,4,2\rangle$. *Find a unit vector that is perpendicular to both* $\boldsymbol{x}$ *and* $\boldsymbol{y}$.

Exercise 12.3.18 *Do the three points* $(1,2,0)$, $(-2,1,1)$, *and* $(0,3,-1)$ *form a right triangle?*

Exercise 12.3.19 *Do the three points* $(1,1,1)$, $(2,3,2)$, *and* $(5,0,-1)$ *form a right triangle?*

Exercise 12.3.20 *Show that* $|\boldsymbol{v}\cdot\boldsymbol{w}| \le |\boldsymbol{v}||\boldsymbol{w}|$

Exercise 12.3.21 *Let* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *be perpendicular vectors. Use Theorem 12.6 to prove that* $|\boldsymbol{x}|^2 + |\boldsymbol{y}|^2 = |\boldsymbol{x}+\boldsymbol{y}|^2$. *What is this result better known as?*

Exercise 12.3.22 *Prove that the diagonals of a rhombus intersect at right angles.*

Exercise 12.3.23 *Suppose that* $\boldsymbol{z} = |\boldsymbol{x}|\boldsymbol{y} + |\boldsymbol{y}|\boldsymbol{x}$ *where* $\boldsymbol{x}$, $\boldsymbol{y}$, *and* $\boldsymbol{z}$ *are all nonzero vectors. Prove that* $\boldsymbol{z}$ *bisects the angle between* $\boldsymbol{x}$ *and* $\boldsymbol{y}$.

Exercise 12.3.24 *Prove Theorem 12.6.*

12.4 The Cross Product

Suppose we are given two vectors. In many cases it is useful to find a third vector perpendicular to the first two. There are of course an infinite number of such vectors of different lengths. Nevertheless, let us find one. Suppose $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$ and $\mathbf{w} = \langle w_1, w_2, w_3 \rangle$. We want to find a vector $\mathbf{c} = \langle c_1, c_2, c_3 \rangle$ with $\mathbf{c} \cdot \mathbf{v} = \mathbf{c} \cdot \mathbf{w} = 0$, or

$$\begin{aligned} v_1 c_1 + v_2 c_2 + v_3 c_3 &= 0, \\ w_1 c_1 + w_2 c_2 + w_3 c_3 &= 0. \end{aligned}$$

Multiply the first equation by w_3 and the second by v_3 and subtract to get

$$\begin{aligned} w_3 v_1 c_1 + w_3 v_2 c_2 + w_3 v_3 c_3 &= 0 \\ v_3 w_1 c_1 + v_3 w_2 c_2 + v_3 w_3 c_3 &= 0 \\ (v_1 w_3 - w_1 v_3) c_1 + (v_2 w_3 - w_2 v_3) c_2 &= 0 \end{aligned}$$

Of course, this equation in two variables has many solutions; a particularly easy one to see is $c_1 = v_2 w_3 - w_2 v_3$, $c_2 = w_1 v_3 - v_1 w_3$. Substituting back into either of the original equations and solving for c_3 gives $c_3 = v_1 w_2 - w_1 v_2$.

This particular answer to the problem turns out to have some nice properties, and it is dignified with a name: the **cross product**:

$$\mathbf{v} \times \mathbf{w} = \langle v_2 w_3 - w_2 v_3, w_1 v_3 - v_1 w_3, v_1 w_2 - w_1 v_2 \rangle.$$

While there is a nice pattern to this vector, it can be a bit difficult to memorize; here is a convenient mnemonic. The determinant of a two by two matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - cb.$$

This is extended to the determinant of a three by three matrix:

$$\begin{aligned} \begin{vmatrix} x & y & z \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} &= x \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - y \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + z \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} \\ &= x(v_2 w_3 - w_2 v_3) - y(v_1 w_3 - w_1 v_3) + z(v_1 w_2 - w_1 v_2) \\ &= x(v_2 w_3 - w_2 v_3) + y(w_1 v_3 - v_1 w_3) + z(v_1 w_2 - w_1 v_2). \end{aligned}$$

Each of the two by two matrices is formed by deleting the top row and one column of the three by three matrix; the subtraction of the middle term must also be memorized. This is not the place to extol the uses of the determinant; suffice it to say that determinants are extraordinarily useful and important. Here we want to use it merely as a mnemonic device. You will have noticed that the three expressions in parentheses on the last line are precisely the three coordinates of the cross product; replacing x, y, z by $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ gives us

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = (v_2 w_3 - w_2 v_3)\mathbf{i} - (v_1 w_3 - w_1 v_3)\mathbf{j} + (v_1 w_2 - w_1 v_2)\mathbf{k}$$

$$\begin{aligned}
&= (v_2w_3 - w_2v_3)\mathbf{i} + (w_1v_3 - v_1w_3)\mathbf{j} + (v_1w_2 - w_1v_2)\mathbf{k} \\
&= \langle v_2w_3 - w_2v_3, w_1v_3 - v_1w_3, v_1w_2 - w_1v_2\rangle \\
&= \mathbf{v}\times\mathbf{w}.
\end{aligned}$$

Given $\mathbf{v}$ and $\mathbf{w}$, there are typically two possible directions and an infinite number of magnitudes that will give a vector perpendicular to both $\mathbf{v}$ and $\mathbf{w}$. As we have picked a particular one, we should investigate the magnitude and direction.

We know how to compute the magnitude of $\mathbf{v}\times\mathbf{w}$; it's a bit messy but not difficult. It is somewhat easier to work initially with the square of the magnitude, so as to avoid the square root:

$$\begin{aligned}
|\mathbf{v}\times\mathbf{w}|^2 &= (v_2w_3 - w_2v_3)^2 + (w_1v_3 - v_1w_3)^2 + (v_1w_2 - w_1v_2)^2 \\
&= v_2^2w_3^2 - 2v_2w_3w_2v_3 + w_2^2v_3^2 + w_1^2v_3^2 - 2w_1v_3v_1w_3 + v_1^2w_3^2 + v_1^2w_2^2 - 2v_1w_2w_1v_2 + w_1^2v_2^2
\end{aligned}$$

While it is far from obvious, this nasty looking expression can be simplified:

$$\begin{aligned}
|\mathbf{v}\times\mathbf{w}|^2 &= (v_1^2+v_2^2+v_3^2)(w_1^2+w_2^2+w_3^2) - (v_1w_1+v_2w_2+v_3w_3)^2 \\
&= |\mathbf{v}|^2|\mathbf{w}|^2 - (\mathbf{v}\cdot\mathbf{w})^2 \\
&= |\mathbf{v}|^2|\mathbf{w}|^2 - |\mathbf{v}|^2|\mathbf{w}|^2\cos^2\theta \\
&= |\mathbf{v}|^2|\mathbf{w}|^2(1-\cos^2\theta) \\
&= |\mathbf{v}|^2|\mathbf{w}|^2\sin^2\theta \\
|\mathbf{v}\times\mathbf{w}| &= |\mathbf{v}||\mathbf{w}|\sin\theta
\end{aligned}$$

The magnitude of $\mathbf{v}\times\mathbf{w}$ is thus very similar to the dot product. In particular, notice that if $\mathbf{v}$ is parallel to $\mathbf{w}$, the angle between them is zero, so $\sin\theta = 0$, so $|\mathbf{v}\times\mathbf{w}| = 0$, and likewise if they are anti-parallel, $\sin\theta = 0$, and $|\mathbf{v}\times\mathbf{w}| = 0$. Conversely, if $|\mathbf{v}\times\mathbf{w}| = 0$ and $|\mathbf{v}|$ and $|\mathbf{w}|$ are not zero, it must be that $\sin\theta = 0$, so $\mathbf{v}$ is parallel or anti-parallel to $\mathbf{w}$.

Here is a curious fact about this quantity that turns out to be quite useful later on: Given two vectors, we can put them tail to tail and form a parallelogram, as in Figure 12.10. The height of the parallelogram, h, is $|\mathbf{v}|\sin\theta$, and the base is $|\mathbf{w}|$, so the area of the parallelogram is $|\mathbf{v}||\mathbf{w}|\sin\theta$, exactly the magnitude of $|\mathbf{v}\times\mathbf{w}|$.

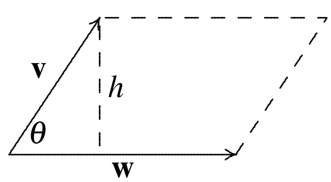


Figure 12.10: A parallelogram.

What about the direction of the cross product? Remarkably, there is a simple rule that describes the direction. Let's look at a simple example: Let $\mathbf{v} = \langle a,0,0\rangle$, $\mathbf{w} = \langle b,c,0\rangle$. If the vectors are placed with tails at the origin, $\mathbf{v}$ lies along the x-axis and $\mathbf{w}$ lies in the x-y plane, so we know the cross product will point either up or down. The cross product is

$$\mathbf{v}\times\mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a & 0 & 0 \\ b & c & 0 \end{vmatrix} = \langle 0,0,ac\rangle.$$

As predicted, this is a vector pointing up or down, depending on the sign of ac. Suppose that $a > 0$, so the sign depends only on c: if $c > 0$, $ac > 0$ and the vector points up; if $c < 0$, the vector points down. On the other hand, if $a < 0$ and $c > 0$, the vector points down, while if $a < 0$ and $c < 0$, the vector points up. Here is how to interpret these facts with a single rule: Imagine rotating vector $\mathbf{v}$ until it points in the same direction as $\mathbf{w}$; there are two ways to do this—use the rotation that goes through the smaller angle. If $a > 0$ and $c > 0$, or $a < 0$ and $c < 0$, the rotation will be counter-clockwise when viewed from above; in the other two cases, $\mathbf{v}$ must be rotated clockwise to reach $\mathbf{w}$. The rule is: counter-clockwise means up, clockwise means down. If $\mathbf{v}$ and $\mathbf{w}$ are any vectors in the x-y plane, the same rule applies—$\mathbf{v}$ need not be parallel to the x-axis.

Although it is somewhat difficult computationally to see how this plays out for any two starting vectors, the rule is essentially the same. Place $\mathbf{v}$ and $\mathbf{w}$ tail to tail. The plane in which $\mathbf{v}$ and $\mathbf{w}$ lie may be viewed from two sides; view it from the side for which $\mathbf{v}$ must rotate counter-clockwise to reach $\mathbf{w}$; then the vector $\mathbf{v} \times \mathbf{w}$ points toward you.

This rule is usually called the **right hand rule**. Imagine placing the heel of your right hand at the point where the tails are joined, so that your slightly curled fingers indicate the direction of rotation from $\mathbf{v}$ to $\mathbf{w}$. Then your thumb points in the direction of the cross product $\mathbf{v} \times \mathbf{w}$.

One immediate consequence of these facts is that $\mathbf{v} \times \mathbf{w} \neq \mathbf{w} \times \mathbf{v}$, because the two cross products point in the opposite direction. On the other hand, since

$$|\mathbf{v} \times \mathbf{w}| = |\mathbf{v}||\,\mathbf{w}|\sin\theta = |\mathbf{w}||\mathbf{v}|\sin\theta = |\mathbf{w} \times \mathbf{v}|,$$

the lengths of the two cross products are equal, so we know that $\mathbf{v} \times \mathbf{w} = -(\mathbf{w} \times \mathbf{v})$.

The cross product has some familiar-looking properties that will be useful later, so we list them here. As with the dot product, these can be proved by performing the appropriate calculations on coordinates, after which we may sometimes avoid such calculations by using the properties.

Theorem 12.7: Cross Product Properties

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors and a is a real number, then

1. $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$
2. $(\mathbf{v} + \mathbf{w}) \times \mathbf{u} = \mathbf{v} \times \mathbf{u} + \mathbf{w} \times \mathbf{u}$
3. $(a\mathbf{u}) \times \mathbf{v} = a(\mathbf{u} \times \mathbf{v}) = \mathbf{u} \times (a\mathbf{v})$
4. $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$
5. $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$

Exercises for 12.4

Exercise 12.4.1 *Find the cross product of* $\langle 1,1,1\rangle$ *and* $\langle 1,2,3\rangle$.

Exercise 12.4.2 *Find the cross product of* $\langle 1,0,2\rangle$ *and* $\langle -1,-2,4\rangle$.

Exercise 12.4.3 *Find the cross product of* $\langle -2,1,3\rangle$ *and* $\langle 5,2,-1\rangle$.

Exercise 12.4.4 *Find the cross product of* $\langle 1,0,0\rangle$ *and* $\langle 0,0,1\rangle$.

Exercise 12.4.5 *Two vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are separated by an angle of* $\pi/6$, *and* $|\boldsymbol{u}| = 2$ *and* $|\boldsymbol{v}| = 3$. *Find* $|\boldsymbol{u}\times\boldsymbol{v}|$.

Exercise 12.4.6 *Two vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are separated by an angle of* $\pi/4$, *and* $|\boldsymbol{u}| = 3$ *and* $|\boldsymbol{v}| = 7$. *Find* $|\boldsymbol{u}\times\boldsymbol{v}|$.

Exercise 12.4.7 *Find the area of the parallelogram with vertices* $(0,0)$, $(1,2)$, $(3,7)$, *and* $(2,5)$.

Exercise 12.4.8 *Find and explain the value of* $(\boldsymbol{i}\times\boldsymbol{j})\times\boldsymbol{k}$ *and* $(\boldsymbol{i}+\boldsymbol{j})\times(\boldsymbol{i}-\boldsymbol{j})$.

Exercise 12.4.9 *Prove that for all vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$, $(\boldsymbol{u}\times\boldsymbol{v})\cdot\boldsymbol{v} = 0$.

Exercise 12.4.10 *Prove Theorem 12.7.*

Exercise 12.4.11 *Define the triple product of three vectors,* $\boldsymbol{x}$, $\boldsymbol{y}$, *and* $\boldsymbol{z}$, *to be the scalar* $\boldsymbol{x}\cdot(\boldsymbol{y}\times\boldsymbol{z})$. *Show that three vectors lie in the same plane if and only if their triple product is zero. Verify that* $\langle 1,5,-2\rangle$, $\langle 4,3,0\rangle$ *and* $\langle 6,13,-4\rangle$ *all lie in the same plane.*

12.5 Lines and Planes

Lines and planes are perhaps the simplest of curves and surfaces in three dimensional space. They also will prove important as we seek to understand more complicated curves and surfaces.

You may recall that the equation of a line in two dimensions is $ax+by=c$; it is reasonable to expect that a line in three dimensions is given by $ax+by+cz=d$. However it turns out that this is the equation of a plane. We will turn our attention to a study of planes and return to consider lines later in this section.

A plane does not have an obvious "direction" as does a line. It is possible to associate a plane with a direction in a very useful way, however: there are exactly two directions perpendicular to a plane. Any vector with one of these two directions is called **normal** to the plane. While there are many normal vectors to a given plane, they are all parallel or anti-parallel to each other.

Suppose two points (v_1,v_2,v_3) and (w_1,w_2,w_3) are in a plane; then the vector $\langle w_1-v_1,w_2-v_2,w_3-v_3\rangle$ is parallel to the plane. In particular, if this vector is placed with its tail at (v_1,v_2,v_3) then its head is at (w_1,w_2,w_3) and it lies in the plane. As a result, any vector perpendicular to the plane is perpendicular to $\langle w_1-v_1,w_2-v_2,w_3-v_3\rangle$. In fact, it is easy to see that the plane consists of *precisely* those points (w_1,w_2,w_3) for which $\langle w_1-v_1,w_2-v_2,w_3-v_3\rangle$ is perpendicular to a normal to the plane, as indicated in Figure 12.11. Turning this around, suppose we know that $\langle a,b,c\rangle$ is normal to a plane containing the point (v_1,v_2,v_3). Then (x,y,z) is in the plane if and only if $\langle a,b,c\rangle$ is perpendicular to $\langle x-v_1,y-v_2,z-v_3\rangle$. In turn, we know that this is true precisely when $\langle a,b,c\rangle \cdot \langle x-v_1, y-v_2, z-v_3\rangle = 0$. That is, (x,y,z) is in the plane if and only if

$$\begin{aligned}
\langle a,b,c\rangle \cdot \langle x-v_1, y-v_2, z-v_3\rangle &= 0\\
a(x-v_1)+b(y-v_2)+c(z-v_3) &= 0\\
ax+by+cz-av_1-bv_2-cv_3 &= 0\\
ax+by+cz &= av_1+bv_2+cv_3.
\end{aligned}$$

Working backwards, note that if (x,y,z) is a point satisfying $ax+by+cz=d$ then

$$\begin{aligned}
ax+by+cz &= d\\
ax+by+cz-d &= 0\\
a(x-d/a)+b(y-0)+c(z-0) &= 0\\
\langle a,b,c\rangle \cdot \langle x-d/a, y, z\rangle &= 0.
\end{aligned}$$

Namely, $\langle a,b,c\rangle$ is perpendicular to the vector with tail at $(d/a,0,0)$ and head at (x,y,z). This means that the points (x,y,z) that satisfy the equation $ax+by+cz=d$ form a plane perpendicular to $\langle a,b,c\rangle$. (This doesn't work if $a=0$, but in that case we can use b or c in the role of a. That is, either $a(x-0)+b(y-d/b)+c(z-0)=0$ or $a(x-0)+b(y-0)+c(z-d/c)=0$.)

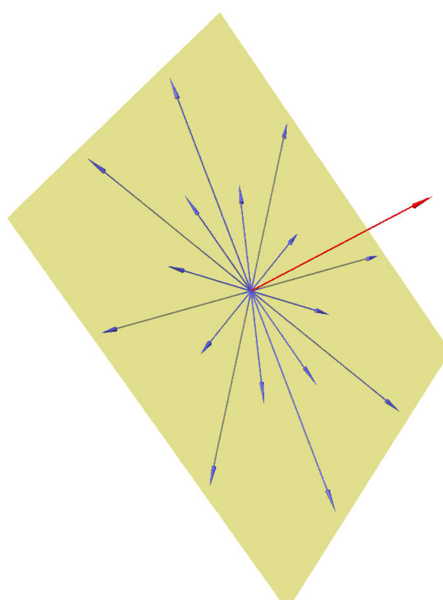


Figure 12.11: A plane defined via vectors perpendicular to a normal.

Thus, given a vector $\langle a,b,c\rangle$ we know that all planes perpendicular to this vector have the form $ax+by+cz=d$, and any surface of this form is a plane perpendicular to $\langle a,b,c\rangle$.

Standard Form of a Plane

Any plane can be written in the form

$$ax + by + cz = d$$

where a, b, c, d are constants and not all a, b, c are zero.
This plane is perpendicular to the vector $\langle a, b, c\rangle$.

Example 12.8: Perpendicular Plane

Find an equation for the plane perpendicular to $\langle 1,2,3\rangle$ *and containing the point* $(5,0,7)$.

Solution. Using the formula above, the plane is $1x + 2y + 3z = d$. To find d we may substitute the known point on the plane to get $5 + 2\cdot 0 + 3\cdot 7 = d$, so $d = 26$. ♣

Example 12.9: Normal Vector

Find a vector normal to the plane $2x - 3y + z = 15$.

Solution. One example is $\langle 2, -3, 1\rangle$. Any vector parallel or anti-parallel to this works as well, so for example $-2\langle 2, -3, 1\rangle = \langle -4, 6, -2\rangle$ is also normal to the plane. ♣

We will frequently need to find an equation for a plane given certain information about the plane. While there may occasionally be slightly shorter ways to get to the desired result, it is always possible, and usually advisable, to use the given information to find a normal to the plane and a point on the plane, and then to find the equation as above.

Example 12.10: Plane Perpendicular

The planes $x - z = 1$ *and* $y + 2z = 3$ *intersect in a line. Find a third plane that contains this line and is perpendicular to the plane* $x + y - 2z = 1$.

Solution. First, we note that two planes are perpendicular if and only if their normal vectors are perpendicular. Thus, we seek a vector $\langle a, b, c\rangle$ that is perpendicular to $\langle 1, 1, -2\rangle$. In addition, since the desired plane is to contain a certain line, $\langle a, b, c\rangle$ must be perpendicular to any vector parallel to this line. Since $\langle a, b, c\rangle$ must be perpendicular to two vectors, we may find it by computing the cross product of the two.

Therefore we need a vector parallel to the line of intersection of the given planes. For this, it suffices to know two points on the line. To find two points on this line, we must find two points that are simultaneously on the two planes, $x - z = 1$ and $y + 2z = 3$. Any point on both planes will satisfy $x - z = 1$ and $y + 2z = 3$. It is easy to find values for x and z satisfying the first, such as $x = 1, z = 0$ and $x = 2, z = 1$. Then we can find corresponding values for y using the second equation, namely $y = 3$ and $y = 1$, so $(1,3,0)$ and $(2,1,1)$ are two such points. They are both on the line of intersection since they are contained in both planes.

Now $\langle 2-1, 1-3, 1-0\rangle = \langle 1, -2, 1\rangle$ is parallel to the line. Finally, we may choose $\langle a, b, c\rangle = \langle 1, 1, -2\rangle \times \langle 1, -2, 1\rangle = \langle -3, -3, -3\rangle$. While this vector will do perfectly well, any vector parallel or anti-parallel to it will work as well. For example we might choose $\langle 1,1,1\rangle$ which is anti-parallel to it, and easier to work with.

Now we know that $\langle 1,1,1\rangle$ is normal to the desired plane and $(2,1,1)$ is a point on the plane. This gives an equation of $(1)x+(1)y+(1)z=d$. Substituting the value of the point into the equation gives $d=4$, and therefore an equation of the plane is $x+y+z=4$. As a quick check, since $(1,3,0)$ is also on the line, it should be on the plane; since $1+3+0=4$, we see that this is indeed the case.

Note that had we used $\langle -3,-3,-3\rangle$ as the normal, we would have discovered the equation $-3x-3y-3z=-12$. Then we might well have noticed that we could divide both sides by -3 to get the equivalent $x+y+z=4$. ♣

We will now turn our attention to a study of lines. Unfortunately, it turns out to be quite inconvenient to represent a typical line with a single equation; we need to approach lines in a different way.

Unlike a plane, a line in three dimensions does have an obvious direction, namely, the direction of any vector parallel to it. In fact a line can be defined and uniquely identified by providing one point on the line and a vector parallel to the line (in one of two possible directions). That is, the line consists of exactly those points we can reach by starting at the point and going for some distance in the direction of the vector. Let's see how we can translate this into more mathematical language.

Suppose a line contains the point (v_1,v_2,v_3) and is parallel to the vector $\langle a,b,c\rangle$. If we place the vector $\langle v_1,v_2,v_3\rangle$ with its tail at the origin and its head at (v_1,v_2,v_3), and if we place the vector $\langle a,b,c\rangle$ with its tail at (v_1,v_2,v_3), then the head of $\langle a,b,c\rangle$ is at a point on the line. We can get to *any* point on the line by doing the same thing, except using $t\langle a,b,c\rangle$ in place of $\langle a,b,c\rangle$, where t is some real number. Because of the way vector addition works, the point at the head of the vector $t\langle a,b,c\rangle$ is the point at the head of the vector $\langle v_1,v_2,v_3\rangle+t\langle a,b,c\rangle$, namely (v_1+ta,v_2+tb,v_3+tc); see Figure 12.12.

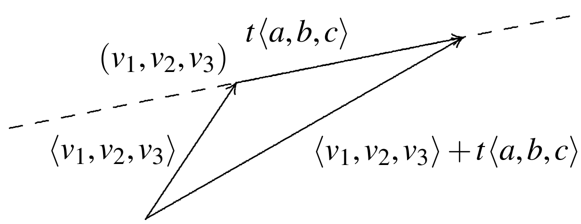


Figure 12.12: Vector form of a line.

In other words, as t runs through all possible real values, the vector $\langle v_1,v_2,v_3\rangle+t\langle a,b,c\rangle$ points to every point on the line when its tail is placed at the origin. It is occasionally useful to use this form of a line even in two dimensions; a vector form for a line in the x-y plane is $\langle v_1,v_2\rangle+t\langle a,b\rangle$, which is the same as $\langle v_1,v_2,0\rangle+t\langle a,b,0\rangle$.

Vector Equation of a Line

An equation for a line passing through point (v_1,v_2,v_3) and parallel to the vector$\langle a,b,c\rangle$ is

$$\langle v_1,v_2,v_3\rangle+t\langle a,b,c\rangle$$

The vector $\langle a,b,c\rangle$ is called the **direction vector** for the line.

Another common way to write this is as a set of **parametric equations**:

$$x = v_1 + ta \qquad y = v_2 + tb \qquad z = v_3 + tc.$$

Parametric Equations of a Line

A line in space can be described as

$$\begin{aligned} x &= v_1 + ta \\ y &= v_2 + tb \\ z &= v_3 + tc \end{aligned}$$

where (v_1, v_2, v_3) is a point on the line and $\langle a, b, c\rangle$ is parallel to the line.

Example 12.11: Vector Expression

Find a vector expression for the line through $(6, 1, -3)$ *and* $(2, 4, 5)$.

Solution. To get a vector parallel to the line we subtract $\langle 6, 1, -3\rangle - \langle 2, 4, 5\rangle = \langle 4, -3, -8\rangle$. The line is then given by $\langle 2, 4, 5\rangle + t\langle 4, -3, -8\rangle$; there are of course many other possibilities, such as $\langle 6, 1, -3\rangle + t\langle 4, -3, -8\rangle$. ♣

Example 12.12: Intersecting Lines

Determine whether the lines $\langle 1, 1, 1\rangle + t\langle 1, 2, -1\rangle$ *and* $\langle 3, 2, 1\rangle + t\langle -1, -5, 3\rangle$ *are parallel, intersect, or neither.*

Solution. In two dimensions, two lines either intersect or are parallel; in three dimensions, lines that do not intersect might not be parallel. In this case, since the direction vectors for the lines are not parallel or anti-parallel we know the lines are not parallel. If they intersect, there must be two values a and b so that $\langle 1, 1, 1\rangle + a\langle 1, 2, -1\rangle = \langle 3, 2, 1\rangle + b\langle -1, -5, 3\rangle$. That is, the following must have a solution:

$$\begin{aligned} 1 + a &= 3 - b \\ 1 + 2a &= 2 - 5b \\ 1 - a &= 1 + 3b \end{aligned}$$

This gives three equations in two unknowns, so there may or may not be a solution in general. In this case, it is easy to discover that $a = 3$ and $b = -1$ satisfies all three equations. Substituting these values into $\langle 1, 1, 1\rangle + a\langle 1, 2, -1\rangle = \langle 3, 2, 1\rangle + b\langle -1, -5, 3\rangle$, the point of intersection is $(4, 7, -2)$. ♣

Example 12.13: Distance from a Point to a Plane

Find the distance from the point $(1, 2, 3)$ *to the plane* $2x - y + 3z = 5$.

Solution. The distance from a point P to a plane is the shortest distance from P to any point on the plane; this is the distance measured from P perpendicular to the plane; see Figure 12.13. This distance is the absolute value of the scalar projection of $\overrightarrow{QP}$ onto a normal vector $\mathbf{n}$, where Q is any point on the plane. It is easy to find a point on the plane, say $(1,0,1)$. Thus the distance is

$$\frac{\langle 0,2,2\rangle\cdot\langle 2,-1,3\rangle}{|\langle 2,-1,3\rangle|}=\frac{4}{\sqrt{14}}.$$

♣

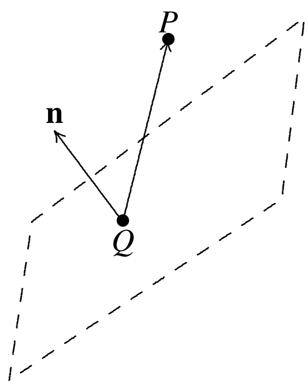


Figure 12.13: Distance from a point to a plane.

Example 12.14: Distance from a Point to a Line

Find the distance from the point $(-1,2,1)$ *to the line* $\langle 1,1,1\rangle+t\langle 2,3,-1\rangle$.

Solution. Again we want the distance measured perpendicular to the line, as indicated in Figure 12.14. The desired distance is

$$|\overrightarrow{QP}|\sin\theta=\frac{|\overrightarrow{QP}\times\mathbf{v}|}{|\mathbf{v}|},$$

where $\mathbf{v}$ is any vector parallel to the line. From the equation of the line, we can use $Q=(1,1,1)$ and $\mathbf{v}=\langle 2,3,-1\rangle$ along with $P=(-1,2,1)$, so the distance is

$$\frac{|\langle -2,1,0\rangle\times\langle 2,3,-1\rangle|}{\sqrt{14}}=\frac{|\langle -1,-2,-8\rangle|}{\sqrt{14}}=\frac{\sqrt{69}}{\sqrt{14}}.$$

♣

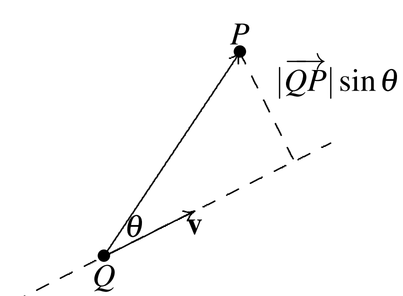


Figure 12.14: Distance from a point to a line.

Exercises for 12.5

Exercise 12.5.1 *Find an equation of the plane containing* $(6,2,1)$ *and perpendicular to* $\langle 1,1,1\rangle$.

Exercise 12.5.2 *Find an equation of the plane containing* $(-1,2,-3)$ *and perpendicular to* $\langle 4,5,-1\rangle$.

Exercise 12.5.3 *Find an equation of the plane containing* $(1,2,-3)$, $(0,1,-2)$ *and* $(1,2,-2)$.

Exercise 12.5.4 *Find an equation of the plane containing* $(1,0,0)$, $(4,2,0)$ *and* $(3,2,1)$.

Exercise 12.5.5 *Find an equation of the plane containing* $(1,0,0)$ *and the line* $\langle 1,0,2\rangle + t\langle 3,2,1\rangle$.

Exercise 12.5.6 *Find an equation of the plane containing the line of intersection of* $x+y+z=1$ *and* $x-y+2z=2$*, and perpendicular to the* x*-*y *plane.*

Exercise 12.5.7 *Find an equation of the line through* $(1,0,3)$ *and* $(1,2,4)$.

Exercise 12.5.8 *Find an equation of the line through* $(1,0,3)$ *and perpendicular to the plane* $x+2y-z=1$.

Exercise 12.5.9 *Find an equation of the line through the origin and perpendicular to the plane* $x+y-z=2$.

Exercise 12.5.10 *Find* a *and* c *so that* $(a,1,c)$ *is on the line through* $(0,2,3)$ *and* $(2,7,5)$.

Exercise 12.5.11 *Explain how to discover the solution in Example 12.12.*

Exercise 12.5.12 *Determine whether the lines* $\langle 1,3,-1\rangle + t\langle 1,1,0\rangle$ *and* $\langle 0,0,0\rangle + t\langle 1,4,5\rangle$ *are parallel, intersect, or neither.*

Exercise 12.5.13 *Determine whether the lines* $\langle 1,0,2\rangle + t\langle -1,-1,2\rangle$ *and* $\langle 4,4,2\rangle + t\langle 2,2,-4\rangle$ *are parallel, intersect, or neither.*

Exercise 12.5.14 *Determine whether the lines* $\langle 1,2,-1\rangle + t\langle 1,2,3\rangle$ *and* $\langle 1,0,1\rangle + t\langle 2/3,2,4/3\rangle$ *are parallel, intersect, or neither.*

Exercise 12.5.15 *Determine whether the lines* $\langle 1,1,2\rangle + t\langle 1,2,-3\rangle$ *and* $\langle 2,3,-1\rangle + t\langle 2,4,-6\rangle$ *are parallel, intersect, or neither.*

Exercise 12.5.16 *Find a unit normal vector to each of the coordinate planes.*

Exercise 12.5.17 *Show that* $\langle 2,1,3\rangle + t\langle 1,1,2\rangle$ *and* $\langle 3,2,5\rangle + s\langle 2,2,4\rangle$ *are the same line.*

Exercise 12.5.18 *Give a prose description for each of the following processes:*

(a) *Given two distinct points, find the line that goes through them.*

(b) *Given three points (not all on the same line), find the plane that goes through them. Why do we need the caveat that not all points be on the same line?*

(c) *Given a line and a point not on the line, find the plane that contains them both.*

(d) *Given a plane and a point not on the plane, find the line that is perpendicular to the plane through the given point.*

Exercise 12.5.19 *Find the distance from* $(2,2,2)$ *to* $x+y+z=-1$.

Exercise 12.5.20 *Find the distance from* $(2,-1,-1)$ *to* $2x-3y+z=2$.

Exercise 12.5.21 *Find the distance from* $(2,-1,1)$ *to* $\langle 2,2,0\rangle + t\langle 1,2,3\rangle$.

Exercise 12.5.22 *Find the distance from* $(1,0,1)$ *to* $\langle 3,2,1\rangle + t\langle 2,-1,-2\rangle$.

Exercise 12.5.23 *Find the cosine of the angle between the planes* $x+y+z=2$ *and* $x+2y+3z=8$.

Exercise 12.5.24 *Find the cosine of the angle between the planes* $x-y+2z=2$ *and* $3x-2y+z=5$.

12.6 Other Coordinate Systems

Coordinate systems are tools that let us use algebraic methods to understand geometry. While the **rectangular** (also called **Cartesian**) coordinates that we have been discussing are the most common, some problems are easier to analyze in alternate coordinate systems.

A coordinate system is a scheme that allows us to identify any point in the plane or in three-dimensional space by a set of numbers. In rectangular coordinates these numbers are interpreted, roughly speaking, as the lengths of the sides of a rectangular "box."

In two dimensions you may already be familiar with an alternative, called **polar coordinates**. In this system, each point in the plane is identified by a pair of numbers (r,θ). The number θ measures the counter-clockwise angle between the positive x-axis and a vector with tail at the origin and head at the point, as shown in Figure 12.15; the number r measures the distance from the origin to the point. Either of these may be negative; a negative θ indicates the angle is measured clockwise from the positive x-axis instead of counter-clockwise, and a negative r indicates the point at distance $|r|$ in the opposite of the direction given by θ.

The relationship between polar and rectangular coordinates is given by

$$
\begin{aligned}
x &= r\cos\theta \\
y &= r\sin\theta
\end{aligned}
$$

and

$$r = \sqrt{x^2 + y^2}$$
$$\tan\theta = \frac{y}{x}$$

Example 12.15: Rectangular to Polar Coordinates

Convert the point $(x,y) = (1,\sqrt{3})$ into polar coordinates.

Solution. First calculate r:

$$r = \sqrt{x^2+y^2} = \sqrt{1+3} = 2$$

Now find θ such that $\tan\theta = \frac{\sqrt{3}}{1}$. The required θ is $\frac{\pi}{3}$.

The polar coordinates are $(2, \frac{\pi}{3})$. ♣

Figure 12.15 also shows the point with rectangular coordinates $(1,\sqrt{3})$ and polar coordinates $(2,\pi/3)$, 2 units from the origin and $\pi/3$ radians from the positive x-axis.

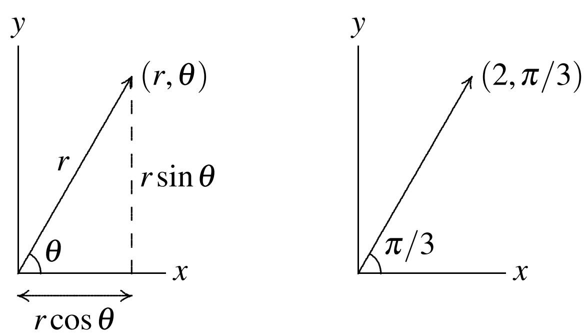


Figure 12.15: Polar coordinates: the general case and the point with rectangular coordinates $(1,\sqrt{3})$.

We can extend polar coordinates to three dimensions simply by adding a z coordinate; this is called **cylindrical coordinates**. Each point in three-dimensional space is represented by three coordinates (r,θ,z) in the obvious way: this point is z units above or below the point (r,θ) in the x-y plane, as shown in Figure 12.16. The point with rectangular coordinates $(1,\sqrt{3},3)$ and cylindrical coordinates $(2,\pi/3,3)$ is also indicated in Figure 12.16.

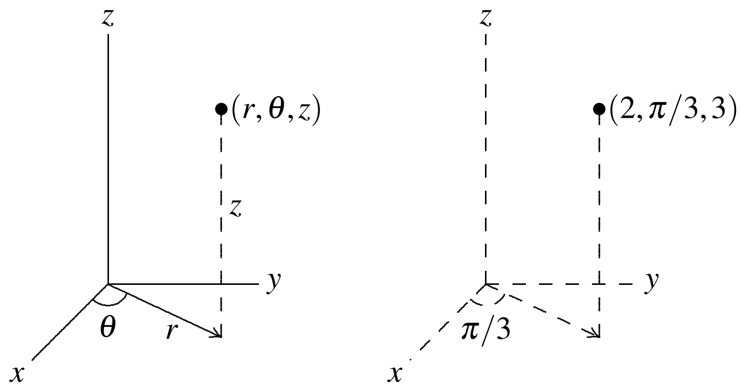


Figure 12.16: Cylindrical coordinates: the general case and the point with rectangular coordinates $(1,\sqrt{3},3)$.

Some figures with relatively complicated equations in rectangular coordinates will be represented by simpler equations in cylindrical coordinates. For example, the cylinder in Figure 12.17 has equation $x^2+y^2=4$ in rectangular coordinates, but equation $r=2$ in cylindrical coordinates.

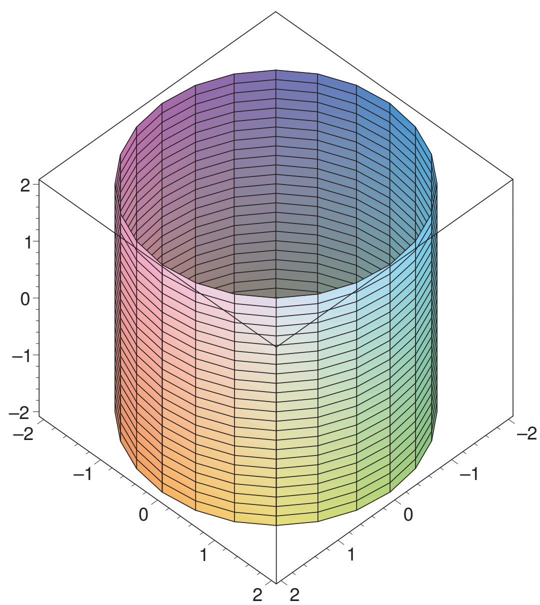


Figure 12.17: The cylinder $r=2$.

Given a point (r,θ) in polar coordinates, it is easy to see (as in Figure 12.15) that the rectangular coordinates of the same point are $(r\cos\theta, r\sin\theta)$, and so the point (r,θ,z) in cylindrical coordinates is $(r\cos\theta, r\sin\theta, z)$ in rectangular coordinates. This means it is usually easy to convert any equation from rectangular to cylindrical coordinates: simply substitute

$$x = r\cos\theta$$
$$y = r\sin\theta$$

and leave z alone. For example, starting with $x^2+y^2=4$ and substituting $x=r\cos\theta$, $y=r\sin\theta$ gives

$$r^2\cos^2\theta + r^2\sin^2\theta = 4$$

$$
\begin{aligned}
r^2(\cos^2\theta + \sin^2\theta) &= 4 \\
r^2 &= 4 \\
r &= 2.
\end{aligned}
$$

Of course, it's easy to see directly that this defines a cylinder as mentioned above.

Cylindrical coordinates are an obvious extension of polar coordinates to three dimensions, but the use of the z coordinate means they are not as closely analogous to polar coordinates as another standard coordinate system. In polar coordinates, we identify a point by a direction and distance from the origin; in three dimensions we can do the same thing, in a variety of ways. The question is: how do we represent a direction? One way is to give the angle of rotation, θ, from the positive x axis, just as in cylindrical coordinates, and also an angle of rotation, ϕ, from the positive z axis. Roughly speaking, θ is like longitude and ϕ is like latitude. (Earth longitude is measured as a positive or negative angle from the prime meridian, and is always between 0 and 180 degrees, east or west; θ can be any positive or negative angle, and we use radians except in informal circumstances. Earth latitude is measured north or south from the equator; ϕ is measured from the north pole down.) This system is called **spherical coordinates**; the coordinates are listed in the order (ρ, θ, ϕ), where ρ is the distance from the origin, and like r in polar and cylindrical coordinates it may be negative. The general case and an example are pictured in Figure 12.18; the length marked r is the r of cylindrical coordinates.

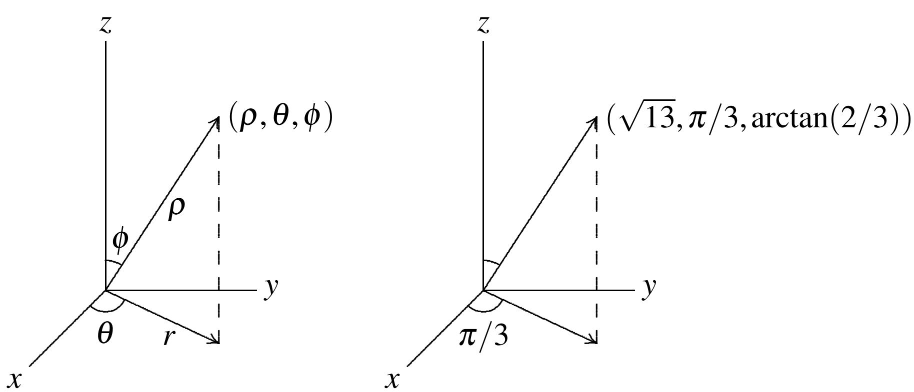


Figure 12.18: Spherical coordinates: the general case and the point with rectangular coordinates $(1, \sqrt{3}, 3)$.

As with cylindrical coordinates, we can easily convert equations in rectangular coordinates to the equivalent in spherical coordinates, though it is a bit more difficult to discover the proper substitutions. Figure 12.19 shows the typical point in spherical coordinates from Figure 12.18, viewed now so that the arrow marked r in the original graph appears as the horizontal "axis" in the left hand graph. From this diagram it is easy to see that the z coordinate is $\rho\cos\phi$, and that $r = \rho\sin\phi$, as shown. Thus, in converting from rectangular to spherical coordinates we will replace z by $\rho\cos\phi$. To see the substitutions for x and y we now view the same point from above, as shown in the right hand graph. The hypotenuse of the triangle in the right hand graph is $r = \rho\sin\phi$, so the sides of the triangle, as shown, are $x = r\cos\theta = \rho\sin\phi\cos\theta$ and $y = r\sin\theta = \rho\sin\phi\sin\theta$. Therefore to convert from rectangular to spherical coordinates, we make these substitutions:

$$
\begin{aligned}
x &= \rho\sin\phi\cos\theta \\
y &= \rho\sin\phi\sin\theta
\end{aligned}
$$

$$z = \rho \cos \phi.$$

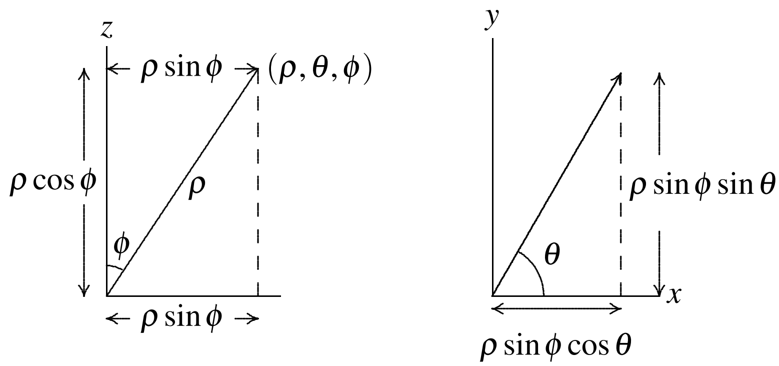


Figure 12.19: Converting from rectangular to spherical coordinates.

As the cylinder had a simple equation in cylindrical coordinates, so does the sphere in spherical coordinates.

Example 12.16:

Find an equation for the sphere of radius 2 *in spherical coordinates.*

Solution. If we start with the Cartesian equation of the sphere and substitute, we get the spherical equation:

$$\begin{aligned} x^2 + y^2 + z^2 &= 2^2 \\ \rho^2 \sin^2 \phi \cos^2 \theta + \rho^2 \sin^2 \phi \sin^2 \theta + \rho^2 \cos^2 \phi &= 2^2 \\ \rho^2 \sin^2 \phi (\cos^2 \theta + \sin^2 \theta) + \rho^2 \cos^2 \phi &= 2^2 \\ \rho^2 \sin^2 \phi + \rho^2 \cos^2 \phi &= 2^2 \\ \rho^2 (\sin^2 \phi + \cos^2 \phi) &= 2^2 \\ \rho^2 &= 2^2 \\ \rho &= 2 \end{aligned}$$

Therefore, in spherical coordinates, a sphere of radius 2 is expressed $\rho = 2$. ♣

Although not as simple as with cylindrical coordinates, we can use spherical coordinates to describe the equation of a cylinder.

Example 12.17: Cylinder Equation in Spherical Coordinates

Find an equation for the cylinder $x^2 + y^2 = 4$ *in spherical coordinates.*

Solution. Proceeding as in the previous example:

$$x^2 + y^2 = 4$$

$$\begin{aligned}\rho^2\sin^2\phi\cos^2\theta+\rho^2\sin^2\phi\sin^2\theta&=4\\ \rho^2\sin^2\phi(\cos^2\theta+\sin^2\theta)&=4\\ \rho^2\sin^2\phi&=4\\ \rho\sin\phi&=2\\ \rho&=\frac{2}{\sin\phi}\end{aligned}$$

♣

Exercises for 12.6

Exercise 12.6.1 *Convert the following points in rectangular coordinates to cylindrical and spherical coordinates:*

(a) $(1,1,1)$

(b) $(7,-7,5)$

(c) $(\cos(1),\sin(1),1)$

(d) $(0,0,-\pi)$

Exercise 12.6.2 *Find an equation for the sphere* $x^2+y^2+z^2=4$ *in cylindrical coordinates.*

Exercise 12.6.3 *Find an equation for the y-z plane in cylindrical coordinates.*

Exercise 12.6.4 *Find an equation equivalent to* $x^2+y^2+2z^2+2z-5=0$ *in cylindrical coordinates.*

Exercise 12.6.5 *Suppose the curve* $z=e^{-x^2}$ *in the x-z plane is rotated around the z axis. Find an equation for the resulting surface in cylindrical coordinates.*

Exercise 12.6.6 *Suppose the curve* $z=x$ *in the x-z plane is rotated around the z axis. Find an equation for the resulting surface in cylindrical coordinates.*

Exercise 12.6.7 *Find an equation for the plane* $y=0$ *in spherical coordinates.*

Exercise 12.6.8 *Find an equation for the plane* $z=1$ *in spherical coordinates.*

Exercise 12.6.9 *Find an equation for the sphere with radius 1 and center at* $(0,1,0)$ *in spherical coordinates.*

Exercise 12.6.10 *Find an equation for the cylinder* $x^2+y^2=4$ *in spherical coordinates.*

Exercise 12.6.11 *Suppose the curve $z = x$ in the x-z plane is rotated around the z axis. Find an equation for the resulting surface in spherical coordinates.*

Exercise 12.6.12 *Plot the polar equations $r = \sin(\theta)$ and $r = \cos(\theta)$ and comment on their similarities. (If you get stuck on how to plot these, you can multiply both sides of each equation by r and convert back to rectangular coordinates).*

Exercise 12.6.13 *Extend Exercises 12.6.6 and 12.6.11 by rotating the curve $z = mx$ around the z axis and converting to both cylindrical and spherical coordinates.*

Exercise 12.6.14 *Convert the spherical formula $\rho = \sin\theta\sin\phi$ to rectangular coordinates and describe the surface defined by the formula (Hint: Multiply both sides by ρ.)*

Exercise 12.6.15 *We can describe points in the first octant by $x > 0$, $y > 0$ and $z > 0$. Give similar inequalities for the first octant in cylindrical and spherical coordinates.*

13. Partial Differentiation

13.1 Functions of Several Variables

In single-variable calculus we were concerned with functions that map the real numbers $\mathbb{R}$ to $\mathbb{R}$, sometimes called "real functions of one variable", meaning the "input" is a single real number and the "output" is likewise a single real number. Now we turn to functions of several variables, where several input variables are mapped to one value: functions $f\colon \mathbb{R}^n \to \mathbb{R}$. We will deal primarily with $n = 2$ and to a lesser extent $n = 3$; in fact many of the techniques we discuss can be applied to larger values of n as well.

A function $f\colon \mathbb{R}^2 \to \mathbb{R}$ maps a pair of values (x, y) to a single real number. The three-dimensional coordinate system we have already used is a convenient way to visualize such functions: above each point (x, y) in the x-y plane we graph the point (x, y, z), where of course $z = f(x, y)$.

Example 13.1: Plane

Describe the function $f(x, y) = 3x + 4y - 5$.

Solution. Writing this as $z = 3x + 4y - 5$ and then $3x + 4y - z = 5$ we recognize the equation of a plane. In the form $f(x, y) = 3x + 4y - 5$ the emphasis has shifted: we now think of x and y as independent variables and z as a variable dependent on them, but the geometry is unchanged. ♣

Example 13.2: Sphere

Describe the equation $x^2 + y^2 + z^2 = 4$.

Solution. We have seen that $x^2 + y^2 + z^2 = 4$ represents a sphere of radius 2. We cannot write this in the form $f(x, y)$, since for each x and y in the disk $x^2 + y^2 < 4$ there are two corresponding points on the sphere. As with the equation of a circle, we can resolve this equation into two functions, $f_1(x, y) = \sqrt{4 - x^2 - y^2}$ and $f_2(x, y) = -\sqrt{4 - x^2 - y^2}$, representing the upper and lower hemispheres, respectively. Each of these is an example of a function with a restricted domain: only certain values of x and y make sense (namely, those for which $x^2 + y^2 \leq 4$) and the graphs of these functions are limited to a small region of the plane. ♣

Example 13.3: Square Root

Describe the function $f(x, y) = \sqrt{x} + \sqrt{y}$.

Solution. This function is defined only when both x and y are non-negative. When $y=0$ we get $f(x,y)=\sqrt{x}$, the familiar square root function in the x-z plane, and when $x=0$ we get the same curve in the y-z plane. Generally speaking, we see that starting from $f(0,0)=0$ this function gets larger in every direction in roughly the same way that the square root function gets larger. For example, if we restrict attention to the line $x=y$, we get $f(x,y)=2\sqrt{x}$ and along the line $y=2x$ we have $f(x,y)=\sqrt{x}+\sqrt{2x}=(1+\sqrt{2})\sqrt{x}$. ♣

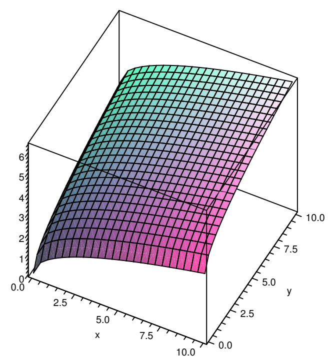


Figure 13.1: $f(x,y)=\sqrt{x}+\sqrt{y}$

A computer program that plots such surfaces can be very useful, as it is often difficult to get a good idea of what they look like. Still, it is valuable to be able to visualize relatively simple surfaces without such aids. As in the previous example, it is often a good idea to examine the function on restricted subsets of the plane, especially lines. It can also be useful to identify those points (x,y) that share a common z-value.

Example 13.4: Elliptic Paraboloid

Describe the graph of $f(x,y)=x^2+y^2$.

Solution. When $x=0$ this becomes $f=y^2$, a parabola in the y-z plane; when $y=0$ we get the "same" parabola $f=x^2$ in the x-z plane.

Finally, picking a value $z=k$, at what points does $f(x,y)=k$? This means $x^2+y^2=k$, which we recognize as the equation of a circle of radius $\sqrt{k}$. So the graph of $f(x,y)$ has parabolic cross-sections, and the same height everywhere on concentric circles with center at the origin. This fits with what we have already discovered. ♣

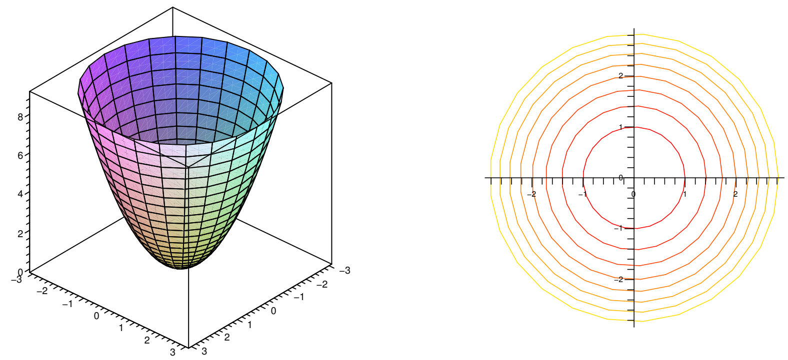


Figure 13.2: $f(x,y) = x^2 + y^2$

As in this example, the points (x,y) such that $f(x,y) = k$ usually form a curve, called a **level curve** of the function. A graph of some level curves can give a good idea of the shape of the surface; it looks much like a topographic map of the surface. In Figure 13.2 both the surface and its associated level curves are shown. Note that, as with a topographic map, the heights corresponding to the level curves are evenly spaced, so that where curves are closer together the surface is steeper.

Functions $f\colon \mathbb{R}^n \to \mathbb{R}$ behave much like functions of two variables; we will on occasion discuss functions of three variables. The principal difficulty with such functions is visualizing them, as they do not "fit" in the three dimensions we are familiar with. For three variables there are various ways to interpret functions that make them easier to understand. For example, $f(x,y,z)$ could represent the temperature at the point (x,y,z), or the pressure, or the strength of a magnetic field. It remains useful to consider those points at which $f(x,y,z) = k$, where k is some constant value. If $f(x,y,z)$ is temperature, the set of points (x,y,z) such that $f(x,y,z) = k$ is the collection of points in space with temperature k; in general this is called a **level set**; for three variables, a level set is typically a surface, called a **level surface**.

Example 13.5: Level Surfaces

Suppose the temperature at (x,y,z) is $T(x,y,z) = e^{-(x^2+y^2+z^2)}$. This function has a maximum value of 1 at the origin, and tends to 0 in all directions. If k is positive and at most 1, the set of points for which $T(x,y,z) = k$ is those points satisfying $x^2+y^2+z^2 = -\ln k$, a sphere centered at the origin. The level surfaces are the concentric spheres centered at the origin.

Exercises for 13.1

Exercise 13.1.1 *Let $f(x,y) = (x-y)^2$. Determine the equations and shapes of the cross-sections when $x = 0$, $y = 0$, $x = y$, and describe the level curves. Use a three-dimensional graphing tool to graph the surface.*

Exercise 13.1.2 *Let $f(x,y) = |x| + |y|$. Determine the equations and shapes of the cross-sections when $x = 0$, $y = 0$, $x = y$, and describe the level curves. Use a three-dimensional graphing tool to graph the surface.*

Exercise 13.1.3 *Let $f(x,y) = e^{-(x^2+y^2)}\sin(x^2 + y^2)$. Determine the equations and shapes of the cross-sections when $x = 0$, $y = 0$, $x = y$, and describe the level curves. Use a three-dimensional graphing tool to graph the surface.*

Exercise 13.1.4 *Let $f(x,y) = \sin(x - y)$. Determine the equations and shapes of the cross-sections when $x = 0$, $y = 0$, $x = y$, and describe the level curves. Use a three-dimensional graphing tool to graph the surface.*

Exercise 13.1.5 *Let $f(x,y) = (x^2 - y^2)^2$. Determine the equations and shapes of the cross-sections when $x = 0$, $y = 0$, $x = y$, and describe the level curves. Use a three-dimensional graphing tool to graph the surface.*

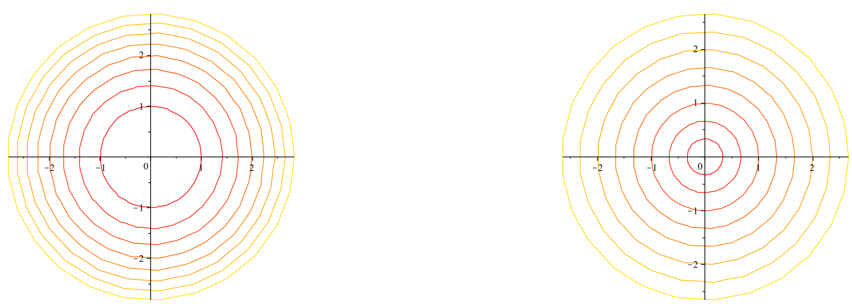
Exercise 13.1.6 *Find the domain of each of the following functions of two variables:*

(a) $\sqrt{9 - x^2} + \sqrt{y^2 - 4}$

(b) $\arcsin(x^2 + y^2 - 2)$

(c) $\sqrt{16 - x^2 - 4y^2}$

Exercise 13.1.7 *Below are two sets of level curves. One is for a cone, one is for a paraboloid. Which is which? Explain.*



13.2 Limits and Continuity

To develop calculus for functions of one variable, we needed to make sense of the concept of a limit, which was used in the definition of a continuous function and the derivative of a function. Limits involving functions of two variables can be considerably more difficult to deal with; fortunately, most of the functions we encounter are fairly easy to understand.

The potential difficulty is largely due to the fact that there are many ways to "approach" a point in the x-y plane. If we want to say that $\lim\limits_{(x,y)\to(a,b)} f(x,y) = L$, we need to capture the idea that as (x,y) gets close to (a,b) then $f(x,y)$ gets close to L. For functions of one variable, $f(x)$, there are only two ways that x can approach a: from the left or right. But there are an infinite number of ways to approach (a,b): along any one of an infinite number of straight lines, or even along a curved path in the x,y-plane. We might hope that it's really not so bad—suppose, for example, that along every possible line through (a,b) the value of $f(x,y)$ gets close to L; surely this means that "$f(x,y)$ approaches L as (x,y) approaches (a,b)". Sadly, no.

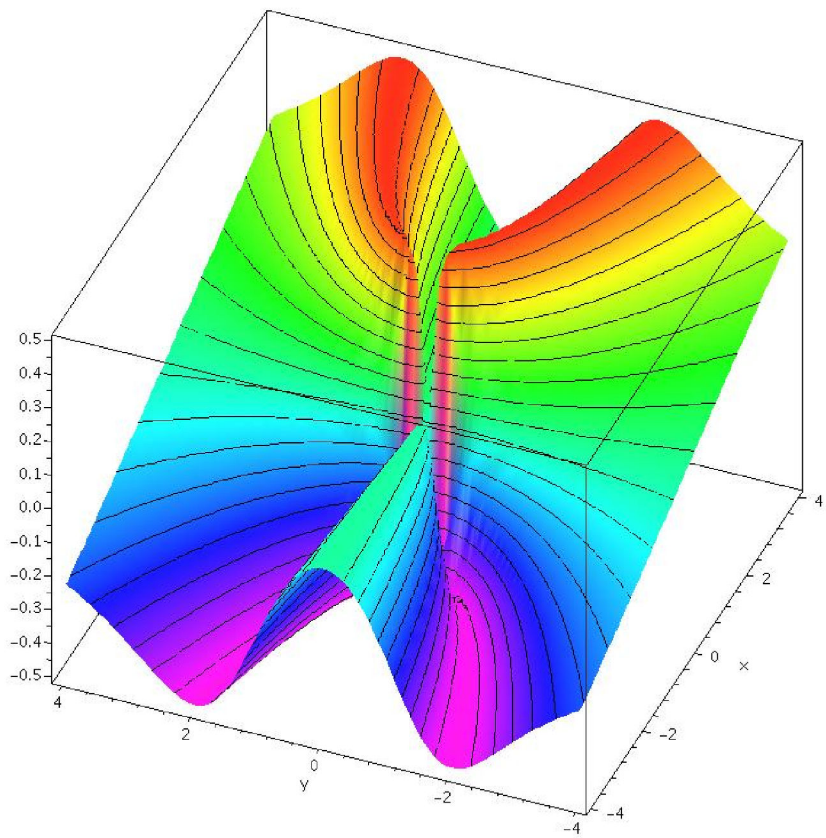


Figure 13.3: $f(x,y) = \dfrac{xy^2}{x^2+y^4}$

Example 13.6: Weird Limit

Analyze $f(x,y) = xy^2/(x^2+y^4)$.

Solution. When $x=0$ or $y=0$, $f(x,y)$ is 0, so the limit of $f(x,y)$ approaching the origin along either the x or y axis is 0. Moreover, along the line $y=mx$, $f(x,y) = m^2x^3/(x^2+m^4x^4)$. As x approaches 0 this expression approaches 0 as well. So along every line through the origin $f(x,y)$ approaches 0. Now suppose we approach the origin along $x=y^2$. Then

$$f(x,y) = \frac{y^2y^2}{y^4+y^4} = \frac{y^4}{2y^4} = \frac{1}{2},$$

so the limit is $1/2$. Looking at Figure 13.3, it is apparent that there is a ridge above $x = y^2$. Approaching the origin along a straight line, we go over the ridge and then drop down toward 0, but approaching along the ridge the height is a constant $1/2$. ♣

Fortunately, we can define the concept of limit without needing to specify how a particular point is approached—indeed, in Definition 3.3, we didn't need the concept of "approach." Roughly, that definition says that when x is close to a then $f(x)$ is close to L; there is no mention of "how" we get close to a. We can adapt that definition to two variables quite easily:

Definition 13.7: Limit of a Multivariate Function

Suppose $f(x,y)$ is a function. We say that

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

if for every $\varepsilon > 0$ there is a $\delta > 0$ so that whenever $0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta$, $|f(x,y) - L| < \varepsilon$.

This says that we can make $|f(x,y) - L| < \varepsilon$, no matter how small ε is, by making the distance from (x,y) to (a,b) "small enough".

Example 13.8: Multivariate Limit

Show that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2} = 0.$

Solution. Suppose $\varepsilon > 0$. Then

$$\left|\frac{3x^2y}{x^2+y^2}\right| = \frac{x^2}{x^2+y^2}3|y|.$$

Note that $x^2/(x^2+y^2) \le 1$ and $|y| = \sqrt{y^2} \le \sqrt{x^2+y^2} < \delta$. So

$$\frac{x^2}{x^2+y^2}3|y| < 1 \cdot 3 \cdot \delta.$$

We want to force this to be less than ε by picking δ "small enough." If we choose $\delta = \varepsilon/3$ then

$$\left|\frac{3x^2y}{x^2+y^2}\right| < 1 \cdot 3 \cdot \frac{\varepsilon}{3} = \varepsilon.$$

♣

Recall that a function $f(x)$ is continuous at $x = a$ if $\lim_{x\to a} f(x) = f(a)$. We can say exactly the same thing about a function of two variables: $f(x,y)$ is continuous at (a,b) if $\displaystyle\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b)$.

The function $f(x,y) = 3x^2y/(x^2+y^2)$ is not continuous at $(0,0)$, because $f(0,0)$ is not defined. However, we know that $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y) = 0$, so we can make a continuous function, by extending the definition of f so that $f(0,0) = 0$. This surface is shown in Figure 13.4.

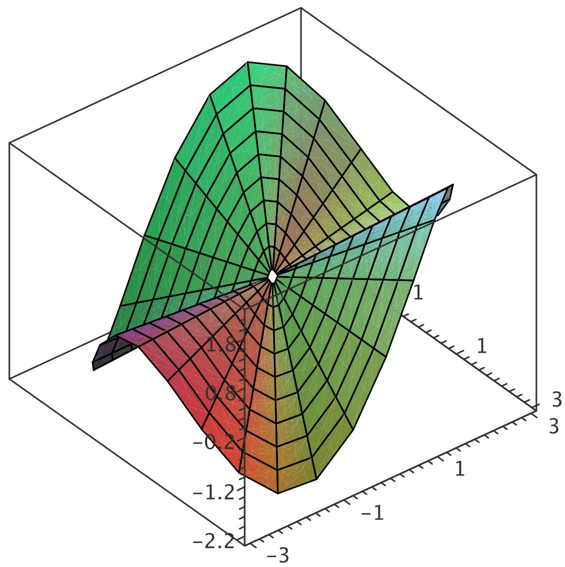


Figure 13.4: $f(x,y) = \dfrac{3x^2y}{x^2+y^2}$

Note that we cannot extend the definition of the function in Example 13.6 to create a continuous function, since the limit does not exist as we approach $(0,0)$.

Fortunately, the functions we will be working with will usually be continuous almost everywhere. As with single variable functions, two classes of common functions are particularly useful and easy to describe. A polynomial in two variables is a sum of terms of the form ax^my^n, where a is a real number and m and n are non-negative integers. A rational function is a quotient of polynomials.

Theorem 13.9: Continuity of Functions

Polynomials are continuous everywhere. Rational functions are continuous everywhere they are defined.

Exercises for 13.2

Determine whether each limit exists. If it does, find the limit and prove that it is the limit; if it does not, explain how you know.

Exercise 13.2.1 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2+y^2}$

Exercise 13.2.2 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$

Exercise 13.2.3 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{2x^2+y^2}$

Exercise 13.2.4 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^4 - y^4}{x^2 + y^2}$

Exercise 13.2.5 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2}$

Exercise 13.2.6 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{2x^2 + y^2}}$

Exercise 13.2.7 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{e^{-x^2-y^2} - 1}{x^2 + y^2}$

Exercise 13.2.8 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^3 + y^3}{x^2 + y^2}$

Exercise 13.2.9 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 + \sin^2 y}{2x^2 + y^2}$

Exercise 13.2.10 $\displaystyle\lim_{(x,y)\to(1,0)} \frac{(x-1)^2 \ln x}{(x-1)^2 + y^2}$

Exercise 13.2.11 $\displaystyle\lim_{(x,y)\to(1,-1)} 3x + 4y$

Exercise 13.2.12 $\displaystyle\lim_{(x,y)\to(0,0)} \frac{4x^2 y}{x^2 + y^2}$

Exercise 13.2.13 *Does the function* $f(x,y) = \dfrac{x-y}{1+x+y}$ *have any discontinuities? What about* $f(x,y) = \dfrac{x-y}{1+x^2+y^2}$*? Explain.*

13.3 Partial Differentiation

The derivative of a function of a single variable tells us how quickly the value of the function changes as the value of the independent variable changes. Intuitively, it tells us how "steep" the graph of the function is. We might wonder if there is a similar idea for graphs of functions of two variables, that is, surfaces. It is not clear that this has a simple answer, nor how we might proceed. We will start with what seem to be very small steps toward the goal. Surprisingly, it turns out that these simple ideas hold the keys to a more general understanding.

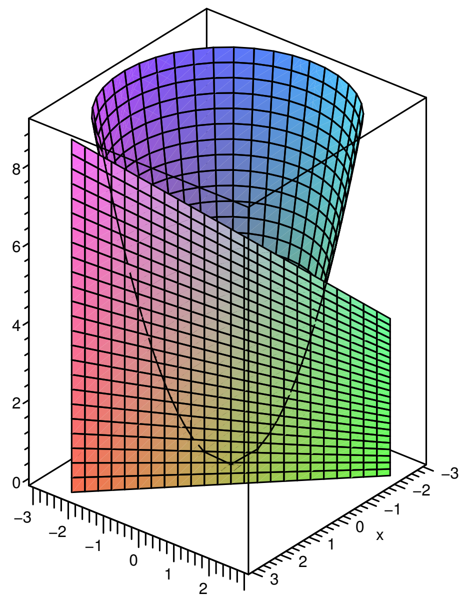


Figure 13.5: The intersection of a plane $x+y=1$ and the surface $f(x,y)=x^2+y^2$

The derivative of a single-variable function $f(x)$ tells us how much $f(x)$ changes as x increases. The obvious analogue for a function of two variables $g(x,y)$ would be something that tells us how quickly $g(x,y)$ increases as x and y increase. However, in most cases this will depend on how quickly x and y are changing relative to each other.

Example 13.10:

Analyze $f(x,y) = y^2$.

Solution. If we look at a point (x,y,y^2) on this surface, the value of a function does not change at all if we fix y and let x increase, but increases like y^2 if we fix x and let y increase. ♣

Now let us consider what happens to $f(x,y)$ when both x and y are increasing, perhaps at different rates. We can think of this as being a movement in a certain direction of a point in the x,y-plane. A point and a direction defines a line in the x,y-plane, and so we are asking how the function changes as we move along this line.

Let us then imagine a plane perpendicular to the x,y-plane that intersects the the x,y-plane along this line. This plane will intersect the surface of f in a curve, so we can just look at the behaviour of this curve in the given plane.

Figure 13.5 shows the plane $x+y=1$, which is the plane perpendicular to the line $x+y=1$ in the x,y-plane. Observe that its intersection with the surface of f is a curve, in fact, a parabola. We will refer to such a curve as the cross-section of the surface above the line in the x,y-plane.

We can now look at the rate of change (or slope) of f in a particular direction by looking at the slope of a curve in a plane — something we already have experience with.

Let's start by looking at some particularly easy lines: Those parallel to the x or y axis. Suppose we are interested in the cross-section of $f(x,y)$ above the line $y=b$. If we substitute b for y in $f(x,y)$, we get a function in one variable, describing the height of the cross-section as a function of x. Because $y=b$ is parallel to the x-axis, if we view it from a vantage point on the negative y-axis, we will see what appears to be simply an ordinary curve in the x-z plane.

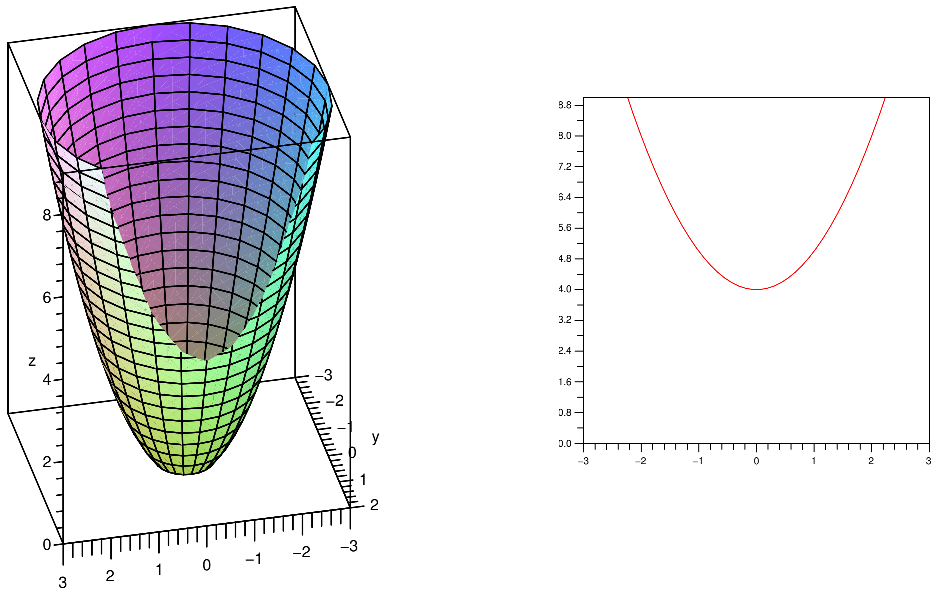


Figure 13.6: $f(x,y) = x^2 + y^2$**, cut by the plane** $y = 2$

Consider again the parabolic surface $f(x,y) = x^2 + y^2$. The cross-section above the line $y = 2$ consists of all points $(x, 2, x^2 + 4)$. Looking at this cross-section we see what appears to be just the curve $f(x) = x^2 + 4$. At any point on the cross-section, $(a, 2, a^2 + 4)$, the slope of the surface *in the direction of the line* $y = 2$ is simply the slope of the curve $f(x) = x^2 + 4$, namely $2x$. Figure 13.6 shows the same parabolic surface as before, but now cut by the plane $y = 2$. The left graph shows the cut-off surface, the right shows just the cross-section.

If, for example, we're interested in the point $(-1, 2, 5)$ on the surface, then the slope in the direction of the line $y = 2$ is $2x = 2(-1) = -2$. This means that starting at $(-1, 2, 5)$ and moving on the surface, above the line $y = 2$, in the direction of increasing x values, the surface goes down; of course moving in the opposite direction, toward decreasing x values, the surface will rise.

If we're interested in some other line $y = k$, there is really no change in the computation. The equation of the cross-section above $y = k$ is $x^2 + k^2$ with derivative $2x$. We can save ourselves the effort, small as it is, of substituting k for y: all we are in effect doing is temporarily assuming that y is some constant. With this assumption, the derivative $\frac{d}{dx}(x^2 + y^2) = 2x$. To emphasize that we are only temporarily assuming y is constant, we use a slightly different notation: $\frac{\partial}{\partial x}(x^2 + y^2) = 2x$; the "$\partial$" reminds us that there are more variables than x, but that only x is being treated as a variable. We read the equation as "the partial derivative of $(x^2 + y^2)$ with respect to x is $2x$." A convenient alternate notation for the partial derivative of $f(x,y)$ with respect to x is is $f_x(x,y)$.

Example 13.11: Partial Derivative with respect to x

Find the partial derivative with respect to x of $x^3 + 3xy$.

Solution. The partial derivative with respect to x of $x^3 + 3xy$ is $3x^2 + 3y$. Note that the partial derivative includes the variable y, unlike the example x^2+y^2. It is somewhat unusual for the partial derivative to depend on a single variable; this example is more typical. ♣

Of course, we can do the same sort of calculation for lines parallel to the y-axis. We temporarily hold x constant, which gives us the equation of the cross-section above a line $x=k$. We can then compute the derivative with respect to y; this will measure the slope of the curve in the y direction.

Example 13.12: Partial Derivative with respect to y

Find the partial derivative with respect to y of $f(x,y)=\sin(xy)+3xy$.

Solution. The partial derivative with respect to y of $f(x,y)=\sin(xy)+3xy$ is

$$f_y(x,y)=\frac{\partial}{\partial y}\sin(xy)+3xy=\cos(xy)\frac{\partial}{\partial y}(xy)+3x=x\cos(xy)+3x.$$

♣

So far, using no new techniques, we have succeeded in measuring the slope of a surface in two quite special directions. For functions of one variable, the derivative is closely linked to the notion of tangent line. For surfaces, the analogous idea is the tangent plane—a plane that just touches a surface at a point, and has the same slope as the surface in all directions. Even though we haven't yet figured out how to compute the slope in all directions, we have enough information to find tangent planes. Suppose we want the plane tangent to a surface at a particular point (a,b,c). If we compute the two partial derivatives of the function for that point, we get enough information to determine two lines tangent to the surface, both through (a,b,c) and both tangent to the surface in their respective directions. These two lines determine a plane, that is, there is exactly one plane containing the two lines: the tangent plane. Figure 13.7 shows (part of) two tangent lines at a point, and the tangent plane containing them.

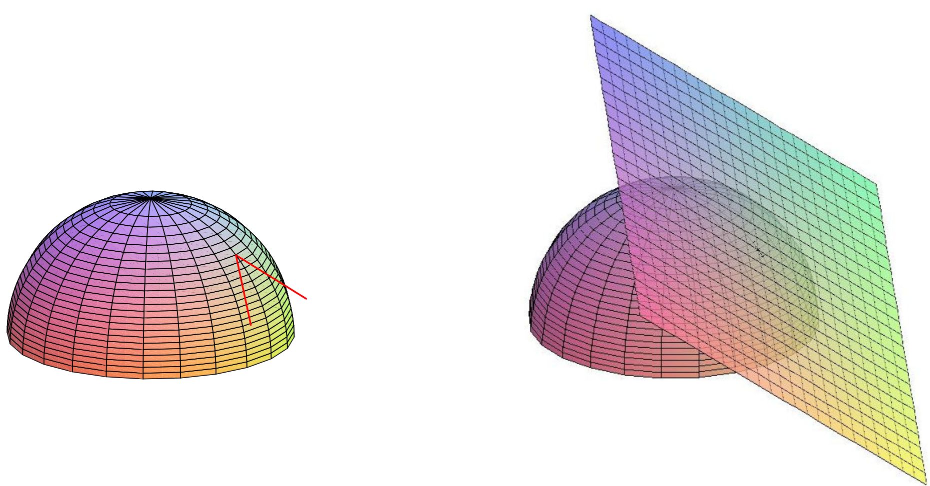


Figure 13.7: Tangent vectors and tangent plane.

How can we discover an equation for this tangent plane? We know a point on the plane, (a,b,c); we need a vector normal to the plane. If we can find two vectors, one parallel to each of the tangent lines we know how to find, then the cross product of these vectors will give the desired normal vector.

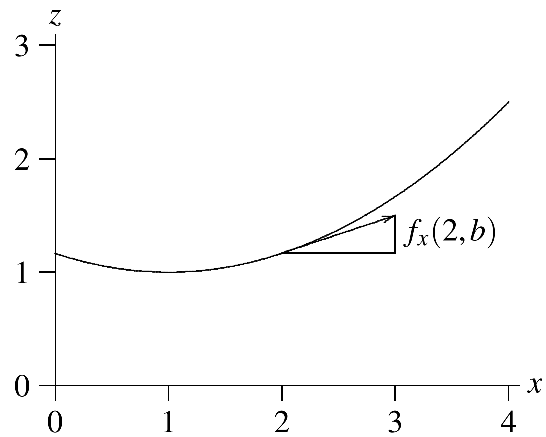


Figure 13.8: A tangent vector.

How can we find vectors parallel to the tangent lines? Consider first the line tangent to the surface above the line $y = b$. A vector $\langle u, v, w\rangle$ parallel to this tangent line must have y component $v = 0$, and we may as well take the x component to be $u = 1$. The ratio of the z component to the x component is the slope of the tangent line, precisely what we know how to compute. The slope of the tangent line is $f_x(a,b)$, so

$$f_x(a,b) = \frac{w}{u} = \frac{w}{1} = w.$$

In other words, a vector parallel to this tangent line is $\langle 1, 0, f_x(a,b)\rangle$, as shown in Figure 13.8. If we repeat the reasoning for the tangent line above $x = a$, we get the vector $\langle 0, 1, f_y(a,b)\rangle$.

Now to find the desired normal vector we compute the cross product, $\langle 0, 1, f_y\rangle \times \langle 1, 0, f_x\rangle = \langle f_x, f_y, -1\rangle$. From our earlier discussion of planes, we can write down the equation we seek: $f_x(a,b)x + f_y(a,b)y - z = k$, and k as usual can be computed by substituting a known point: $f_x(a,b)(a) + f_y(a,b)(b) - c = k$. There are various more-or-less nice ways to write the result:

$$\begin{aligned}
&f_x(a,b)x + f_y(a,b)y - z = f_x(a,b)a + f_y(a,b)b - c\\
&f_x(a,b)x + f_y(a,b)y - f_x(a,b)a - f_y(a,b)b + c = z\\
&f_x(a,b)(x-a) + f_y(a,b)(y-b) + c = z\\
&f_x(a,b)(x-a) + f_y(a,b)(y-b) + f(a,b) = z
\end{aligned}$$

Example 13.13: Tangent Plane to a Sphere

Find the plane tangent to $x^2 + y^2 + z^2 = 4$ at $(1, 1, \sqrt{2})$.

Solution. The point $(1, 1, \sqrt{2})$ is on the upper hemisphere, so we use $f(x,y) = \sqrt{4 - x^2 - y^2}$. Then $f_x(x,y) = -x(4 - x^2 - y^2)^{-1/2}$ and $f_y(x,y) = -y(4 - x^2 - y^2)^{-1/2}$, so $f_x(1,1) = f_y(1,1) = -1/\sqrt{2}$ and the equation of the plane is

$$z = -\frac{1}{\sqrt{2}}(x-1) - \frac{1}{\sqrt{2}}(y-1) + \sqrt{2}.$$

The hemisphere and this tangent plane are pictured in Figure 13.7. ♣

So it appears that to find a tangent plane, we need only find two quite simple ordinary derivatives, namely f_x and f_y. This is true *if the tangent plane exists*. It is, unfortunately, not always the case that if f_x and f_y exist there is a tangent plane. Consider the function $xy^2/(x^2+y^4)$ with $f(0,0)$ defined to be 0, pictured in Figure 13.3. This function has value 0 when $x=0$ or $y=0$. Now it's clear that $f_x(0,0)=f_y(0,0)=0$, because in the x and y directions the surface is simply a horizontal line. But it's also clear from the picture that this surface does not have anything that deserves to be called a tangent plane at the origin, certainly not the x-y plane containing these two tangent lines.

When does a surface have a tangent plane at a particular point? What we really want from a tangent plane, as from a tangent line, is that the plane be a "good" approximation of the surface near the point. Here is how we can make this precise:

Definition 13.14: Tangent Plane

Let $\Delta x = x - x_0$, $\Delta y = y - y_0$, and $\Delta z = z - z_0$ where $z_0 = f(x_0,y_0)$. The function $z = f(x,y)$ is differentiable at (x_0,y_0) if

$$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where both ε_1 and ε_2 approach 0 as (x,y) approaches (x_0,y_0).

This definition takes a bit of absorbing. Let's rewrite the central equation a bit:

$$z = f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) + f(x_0,y_0) + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y. \tag{13.1}$$

The first three terms on the right are the equation of the tangent plane, that is,

$$f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) + f(x_0,y_0)$$

is the z-value of the point on the plane above (x,y). Equation 13.1 says that the z-value of a point on the surface is equal to the z-value of a point on the plane plus a "little bit," namely $\varepsilon_1 \Delta x + \varepsilon_2 \Delta y$. As (x,y) approaches (x_0,y_0), both Δx and Δy approach 0, so this little bit $\varepsilon_1 \Delta x + \varepsilon_2 \Delta y$ also approaches 0, and the z-values on the surface and the plane get close to each other. But that by itself is not very interesting: since the surface and the plane both contain the point (x_0,y_0,z_0), the z values will approach z_0 and hence get close to each other whether the tangent plane is "tangent" to the surface or not. The extra condition in the definition says that as (x,y) approaches (x_0,y_0), the ε values approach 0—this means that $\varepsilon_1 \Delta x + \varepsilon_2 \Delta y$ approaches 0 much, much faster, because $\varepsilon_1 \Delta x$ is much smaller than either ε_1 or Δx. It is this extra condition that makes the plane a tangent plane.

We can see that the extra condition on ε_1 and ε_2 is just what is needed if we look at partial derivatives. Suppose we temporarily fix $y = y_0$, so $\Delta y = 0$. Then the equation from the definition becomes

$$\Delta z = f_x(x_0,y_0)\Delta x + \varepsilon_1 \Delta x$$

or

$$\frac{\Delta z}{\Delta x} = f_x(x_0,y_0) + \varepsilon_1.$$

Now taking the limit of the two sides as Δx approaches 0, the left side turns into the partial derivative of z with respect to x at (x_0,y_0), or in other words $f_x(x_0,y_0)$, and the right side does the same, because as (x,y) approaches (x_0,y_0), ε_1 approaches 0. Essentially the same calculation works for f_y.

Exercises for 13.3

Exercise 13.3.1 *Find f_x and f_y where $f(x,y) = \cos(x^2y) + y^3$.*

Exercise 13.3.2 *Find f_x and f_y where $f(x,y) = \dfrac{xy}{x^2+y}$.*

Exercise 13.3.3 *Find f_x and f_y where $f(x,y) = e^{x^2+y^2}$.*

Exercise 13.3.4 *Find f_x and f_y where $f(x,y) = xy\ln(xy)$.*

Exercise 13.3.5 *Find f_x and f_y where $f(x,y) = \sqrt{1-x^2-y^2}$.*

Exercise 13.3.6 *Find f_x and f_y where $f(x,y) = x\tan(y)$.*

Exercise 13.3.7 *Find f_x and f_y where $f(x,y) = \dfrac{1}{xy}$.*

Exercise 13.3.8 *Find an equation for the plane tangent to $2x^2+3y^2-z^2 = 4$ at $(1,1,-1)$.*

Exercise 13.3.9 *Find an equation for the plane tangent to $f(x,y) = \sin(xy)$ at $(\pi,1/2,1)$.*

Exercise 13.3.10 *Find an equation for the plane tangent to $f(x,y) = x^2+y^3$ at $(3,1,10)$.*

Exercise 13.3.11 *Find an equation for the plane tangent to $f(x,y) = x\ln(xy)$ at $(2,1/2,0)$.*

Exercise 13.3.12 *Find an equation for the line normal to $x^2+4y^2 = 2z$ at $(2,1,4)$.*

Exercise 13.3.13 *Explain in your own words why, when taking a partial derivative of a function of multiple variables, we can treat the variables not being differentiated as constants.*

Exercise 13.3.14 *Consider a differentiable function, $f(x,y)$. Give physical interpretations of the meanings of $f_x(a,b)$ and $f_y(a,b)$ as they relate to the graph of f.*

Exercise 13.3.15 *In much the same way that we used the tangent line to approximate the value of a function from single variable calculus, we can use the tangent plane to approximate a function from multivariable calculus. Consider the tangent plane found in Exercise 13.3.11. Use this plane to approximate $f(1.98,0.4)$.*

Exercise 13.3.16 *The volume of a cylinder is given by $V = \pi r^2h$. Suppose that the current values of r and h are $r = 7$ cm and $h = 3$ cm. Is the volume more sensitive to a small change in radius or the same amount of change in height? Why?*

Exercise 13.3.17 *Suppose that one of your colleagues has calculated the partial derivatives of a given function, and reported to you that $f_x(x,y) = 2x+3y$ and that $f_y(x,y) = 4x+6y$. Do you believe them? Why or why not? If not, what answer might you have accepted for f_y?*

Exercise 13.3.18 *Suppose $f(t)$ and $g(t)$ are single variable differentiable functions. Find $\partial z/\partial x$ and $\partial z/\partial y$ for each of the following two variable functions.*

(a) $z = f(x)g(y)$

(b) $z = f(xy)$

(c) $z = f(x/y)$

13.4 The Chain Rule

Consider the surface $z = x^2y + xy^2$, and suppose that $x = 2+t^4$ and $y = 1-t^3$. We can think of the latter two equations as describing how x and y change relative to, say, time. Then

$$z = x^2y + xy^2 = (2+t^4)^2(1-t^3) + (2+t^4)(1-t^3)^2$$

tells us explicitly how the z coordinate of the corresponding point on the surface depends on t. If we want to know dz/dt we can compute it more or less directly, but it's actually a bit simpler to use product and chain rules:

$$\begin{aligned}\frac{dz}{dt} &= x^2y' + 2xx'y + x2yy' + x'y^2\\ &= (2xy+y^2)x' + (x^2+2xy)y'\\ &= (2(2+t^4)(1-t^3) + (1-t^3)^2)(4t^3) + ((2+t^4)^2 + 2(2+t^4)(1-t^3))(-3t^2)\end{aligned}$$

If we look carefully at the middle step, $dz/dt = (2xy+y^2)x' + (x^2+2xy)y'$, we notice that $2xy+y^2$ is $\partial z/\partial x$, and x^2+2xy is $\partial z/\partial y$. This turns out to be true in general, and gives us a new chain rule:

Theorem 13.15: Multivariate Chain Rule

Suppose that $z = f(x,y)$, f is differentiable, $x = g(t)$, and $y = h(t)$. Assuming that the relevant derivatives exist,

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Proof. If f is differentiable, then

$$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + \varepsilon_1\Delta x + \varepsilon_2\Delta y,$$

where ε_1 and ε_2 approach 0 as (x,y) approaches (x_0,y_0). Then

$$\frac{\Delta z}{\Delta t} = f_x\frac{\Delta x}{\Delta t} + f_y\frac{\Delta y}{\Delta t} + \varepsilon_1\frac{\Delta x}{\Delta t} + \varepsilon_2\frac{\Delta y}{\Delta t}. \tag{13.2}$$

As Δt approaches 0, (x,y) approaches (x_0,y_0) and so

$$\lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t} = \frac{dz}{dt}$$
$$\lim_{\Delta t\to 0}\varepsilon_1\frac{\Delta x}{\Delta t} = 0\cdot\frac{dx}{dt}$$
$$\lim_{\Delta t\to 0}\varepsilon_2\frac{\Delta y}{\Delta t} = 0\cdot\frac{dy}{dt}$$

and so taking the limit of (13.2) as Δt goes to 0 gives

$$\frac{dz}{dt} = f_x\frac{dx}{dt} + f_y\frac{dy}{dt},$$

as desired. ♣

We can write the chain rule in a way that is somewhat closer to the single variable chain rule:

$$\frac{df}{dt} = \langle f_x, f_y\rangle\cdot\langle x', y'\rangle,$$

or (roughly) the derivatives of the outside function "times" the derivatives of the inside functions. Not surprisingly, essentially the same chain rule works for functions of more than two variables, for example, given a function of three variables $f(x,y,z)$, where each of x, y and z is a function of t,

$$\frac{df}{dt} = \langle f_x, f_y, f_z\rangle\cdot\langle x', y', z'\rangle.$$

We can even extend the idea further. Suppose that $f(x,y)$ is a function and $x = g(s,t)$ and $y = h(s,t)$ are functions of two variables s and t. Then f is "really" a function of s and t as well, and

$$\frac{\partial f}{\partial s} = f_x g_s + f_y h_s \qquad \frac{\partial f}{\partial t} = f_x g_t + f_y h_t.$$

The natural extension of this to $f(x,y,z)$ works as well.

Recall that we used the ordinary chain rule to do implicit differentiation. We can do the same with the new chain rule.

Example 13.16: Equation of a Sphere

Find the partial derivative of $x^2+y^2+z^2=4$.

Solution. The equation $x^2+y^2+z^2=4$ defines a sphere, which is not a function of x and y, though it can be thought of as two functions, the top and bottom hemispheres. We can think of z as one of these two functions, so really $z = z(x, y)$, and we can think of x and y as particularly simple functions of x and y, and let $f(x, y, z) = x^2 + y^2 + z^2$. Since $f(x, y, z) = 4$, $\partial f/\partial x = 0$, but using the chain rule:

$$\begin{aligned} 0 = \frac{\partial f}{\partial x} &= f_x \frac{\partial x}{\partial x} + f_y \frac{\partial y}{\partial x} + f_z \frac{\partial z}{\partial x} \\ &= (2x)(1) + (2y)(0) + (2z)\frac{\partial z}{\partial x}, \end{aligned}$$

noting that since y is temporarily held constant its derivative $\partial y/\partial x = 0$. Now we can solve for $\partial z/\partial x$:

$$\frac{\partial z}{\partial x} = -\frac{2x}{2z} = -\frac{x}{z}.$$

In a similar manner we can compute $\partial z/\partial y$. ♣

Exercises for 13.4

Exercise 13.4.1 *Use the chain rule to compute dz/dt for $z = \sin(x^2 + y^2)$, $x = t^2 + 3$, $y = t^3$.*

Exercise 13.4.2 *Use the chain rule to compute dz/dt for $z = x^2 y$, $x = \sin(t)$, $y = t^2 + 1$.*

Exercise 13.4.3 *Use the chain rule to compute $\partial z/\partial s$ and $\partial z/\partial t$ for $z = x^2 y$, $x = \sin(st)$, $y = t^2 + s^2$.*

Exercise 13.4.4 *Use the chain rule to compute $\partial z/\partial s$ and $\partial z/\partial t$ for $z = x^2 y^2$, $x = st$, $y = t^2 - s^2$.*

Exercise 13.4.5 *Use the chain rule to compute $\partial z/\partial x$ and $\partial z/\partial y$ for $2x^2 + 3y^2 - 2z^2 = 9$.*

Exercise 13.4.6 *Use the chain rule to compute $\partial z/\partial x$ and $\partial z/\partial y$ for $2x^2 + y^2 + z^2 = 9$.*

Exercise 13.4.7 *Chemistry students will recognize the* ideal gas law, *given by $PV = nRT$ which relates the Pressure, Volume, and Temperature of n moles of gas. (R is the ideal gas constant). Thus, we can view pressure, volume, and temperature as variables, each one dependent on the other two.*

(a) *If pressure of a gas is increasing at a rate of $0.2Pa/min$ and temperature is increasing at a rate of $1K/min$, how fast is the volume changing?*

(b) *If the volume of a gas is decreasing at a rate of $0.3L/min$ and temperature is increasing at a rate of $.5K/min$, how fast is the pressure changing?*

(c) *If the pressure of a gas is decreasing at a rate of $0.4Pa/min$ and the volume is increasing at a rate of $3L/min$, how fast is the temperature changing?*

Exercise 13.4.8 *Verify the following identity in the case of the ideal gas law:*

$$\frac{\partial P}{\partial V}\frac{\partial V}{\partial T}\frac{\partial T}{\partial P} = -1$$

Exercise 13.4.9 *The previous exercise was a special case of the following fact, which you are to verify here: If $F(x,y,z)$ is a function of 3 variables, and the relation $F(x,y,z) = 0$ defines each of the variables in terms of the other two, namely $x = f(y,z)$, $y = g(x,z)$ and $z = h(x,y)$, then*

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial x} = -1$$

13.5 Directional Derivatives

We still have not answered one of our first questions about the steepness of a surface: starting at a point on a surface given by $f(x,y)$, and walking in a particular direction, how steep is the surface? We are now ready to answer the question.

We already know roughly what has to be done: as shown in Figure 13.5, we extend a line in the x-y plane to a vertical plane, and we then compute the slope of the curve that is the cross-section of the surface in that plane. The major stumbling block is that what appears in this plane to be the horizontal axis, namely the line in the x-y plane, is not an actual axis—we know nothing about the "units" along the axis. Our goal is to make this line into a t axis; then we need formulas to write x and y in terms of this new variable t; then we can write z in terms of t since we know z in terms of x and y; and finally we can simply take the derivative.

So we need to somehow "mark off" units on the line, and we need a convenient way to refer to the line in calculations. It turns out that we can accomplish both by using the vector form of a line. Suppose that $\mathbf{u}$ is a unit vector $\langle u_1, u_2\rangle$ in the direction of interest. A vector equation for the line through (x_0, y_0) in this direction is $\mathbf{v}(t) = \langle u_1 t + x_0, u_2 t + y_0\rangle$. The height of the surface above the point $(u_1 t + x_0, u_2 t + y_0)$ is $g(t) = f(u_1 t + x_0, u_2 t + y_0)$. Because $\mathbf{u}$ is a unit vector, the value of t is precisely the distance along the line from (x_0, y_0) to $(u_1 t + x_0, u_2 t + y_0)$; this means that the line is effectively a t axis, with origin at the point (x_0, y_0), so the slope we seek is

$$\begin{aligned} g'(0) &= \langle f_x(x_0,y_0), f_y(x_0,y_0)\rangle \cdot \langle u_1, u_2\rangle \\ &= \langle f_x, f_y\rangle \cdot \mathbf{u} \\ &= \nabla f \cdot \mathbf{u} \end{aligned}$$

Here we have used the chain rule and the derivatives $\frac{d}{dt}(u_1 t + x_0) = u_1$ and $\frac{d}{dt}(u_2 t + y_0) = u_2$. The vector $\langle f_x, f_y\rangle$ is very useful, so it has its own symbol, ∇f, pronounced "del f"; it is also called the **gradient** of f.

Example 13.17: Slope

Find the slope of $z = x^2 + y^2$ at $(1,2)$ in the direction of the vector $\langle 3,4 \rangle$.

Solution. We first compute the gradient at $(1,2)$: $\nabla f = \langle 2x, 2y \rangle$, which is $\langle 2,4 \rangle$ at $(1,2)$. A unit vector in the desired direction is $\langle 3/5, 4/5 \rangle$, and the desired slope is then $\langle 2,4 \rangle \cdot \langle 3/5, 4/5 \rangle = 6/5 + 16/5 = 22/5$. ♣

Example 13.18: Tangent Vector

Find a tangent vector to $z = x^2 + y^2$ at $(1,2)$ in the direction of the vector $\langle 3,4 \rangle$ and show that it is parallel to the tangent plane at that point.

Solution. Since $\langle 3/5, 4/5 \rangle$ is a unit vector in the desired direction, we can easily expand it to a tangent vector simply by adding the third coordinate computed in the previous example: $\langle 3/5, 4/5, 22/5 \rangle$. To see that this vector is parallel to the tangent plane, we can compute its dot product with a normal to the plane. We know that a normal to the tangent plane is

$$\langle f_x(1,2), f_y(1,2), -1 \rangle = \langle 2, 4, -1 \rangle,$$

and the dot product is $\langle 2,4,-1 \rangle \cdot \langle 3/5, 4/5, 22/5 \rangle = 6/5 + 16/5 - 22/5 = 0$, so the two vectors are perpendicular. (Note that the vector normal to the surface, namely $\langle f_x, f_y, -1 \rangle$, is simply the gradient with a -1 tacked on as the third component.) ♣

The slope of a surface given by $z = f(x,y)$ in the direction of a (two-dimensional) vector $\mathbf{u}$ is called the **directional derivative** of f, written $D_{\mathbf{u}} f$. The directional derivative immediately provides us with some additional information. We know that

$$D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u} = |\nabla f||\mathbf{u}| \cos\theta = |\nabla f| \cos\theta$$

if $\mathbf{u}$ is a unit vector; θ is the angle between ∇f and $\mathbf{u}$. This tells us immediately that the largest value of $D_{\mathbf{u}} f$ occurs when $\cos\theta = 1$, namely, when $\theta = 0$, so ∇f is parallel to $\mathbf{u}$. In other words, the gradient ∇f points in the direction of steepest ascent of the surface, and $|\nabla f|$ is the slope in that direction. Likewise, the smallest value of $D_{\mathbf{u}} f$ occurs when $\cos\theta = -1$, namely, when $\theta = \pi$, so ∇f is anti-parallel to $\mathbf{u}$. In other words, $-\nabla f$ points in the direction of steepest descent of the surface, and $-|\nabla f|$ is the slope in that direction.

Example 13.19: Direction of Steepest Ascent and Descent

Investigate the direction of steepest ascent and descent for $z = x^2 + y^2$.

Solution. The gradient is $\langle 2x, 2y \rangle = 2\langle x, y \rangle$; this is a vector parallel to the vector $\langle x, y \rangle$, so the direction of steepest ascent is directly away from the origin, starting at the point (x,y). The direction of steepest descent is thus directly toward the origin from (x,y). Note that at $(0,0)$ the gradient vector is $\langle 0,0 \rangle$, which has no direction, and it is clear from the plot of this surface that there is a minimum point at the origin, and tangent vectors in all directions are parallel to the x-y plane. ♣

If ∇f is perpendicular to $\mathbf{u}$, $D_{\mathbf{u}}f = |\nabla f|\cos(\pi/2) = 0$, since $\cos(\pi/2) = 0$. This means that in either of the two directions perpendicular to ∇f, the slope of the surface is 0; this implies that a vector in either of these directions is tangent to the level curve at that point. Starting with $\nabla f = \langle f_x, f_y\rangle$, it is easy to find a vector perpendicular to it: either $\langle f_y, -f_x\rangle$ or $\langle -f_y, f_x\rangle$ will work.

If $f(x,y,z)$ is a function of three variables, all the calculations proceed in essentially the same way. The rate at which f changes in a particular direction is $\nabla f \cdot \mathbf{u}$, where now $\nabla f = \langle f_x, f_y, f_z\rangle$ and $\mathbf{u} = \langle u_1, u_2, u_3\rangle$ is a unit vector. Again ∇f points in the direction of maximum rate of increase, $-\nabla f$ points in the direction of maximum rate of decrease, and any vector perpendicular to ∇f is tangent to the level surface $f(x,y,z) = k$ at the point in question. Of course there are no longer just two such vectors; the vectors perpendicular to ∇f describe the tangent plane to the level surface, or in other words ∇f is a normal to the tangent plane.

Example 13.20: Gradient

Suppose the temperature at a point in space is given by $T(x,y,z) = T_0/(1+x^2+y^2+z^2)$; at the origin the temperature in Kelvin is $T_0 > 0$, and it decreases in every direction from there. It might be, for example, that there is a source of heat at the origin, and as we get farther from the source, the temperature decreases. The gradient is

$$\begin{aligned}\nabla T &= \langle \frac{-2T_0x}{(1+x^2+y^2+z^2)^2} + \frac{-2T_0x}{(1+x^2+y^2+z^2)^2} + \frac{-2T_0x}{(1+x^2+y^2+z^2)^2}\rangle \\ &= \frac{-2T_0}{(1+x^2+y^2+z^2)^2}\langle x,y,z\rangle.\end{aligned}$$

The gradient points directly at the origin from the point (x,y,z)—by moving directly toward the heat source, we increase the temperature as quickly as possible.

Example 13.21: Tangent Plane

Find the points on the surface defined by $x^2+2y^2+3z^2 = 1$ where the tangent plane is parallel to the plane defined by $3x-y+3z = 1$.

Solution. Two planes are parallel if their normals are parallel or anti-parallel, so we want to find the points on the surface with normal parallel or anti-parallel to $\langle 3,-1,3\rangle$. Let $f = x^2+2y^2+3z^2$; the gradient of f is normal to the level surface at every point, so we are looking for a gradient parallel or anti-parallel to $\langle 3,-1,3\rangle$. The gradient is $\langle 2x,4y,6z\rangle$; if it is parallel or anti-parallel to $\langle 3,-1,3\rangle$, then

$$\langle 2x,4y,6z\rangle = k\langle 3,-1,3\rangle$$

for some k. This means we need a solution to the equations

$$2x = 3k \qquad 4y = -k \qquad 6z = 3k$$

but this is three equations in four unknowns—we need another equation. What we haven't used so far is that the points we seek are on the surface $x^2+2y^2+3z^2 = 1$; this is the fourth equation. If we solve the first three equations for x, y, and z and substitute into the fourth equation we get

$$\begin{aligned} 1 &= \left(\frac{3k}{2}\right)^2 + 2\left(\frac{-k}{4}\right)^2 + 3\left(\frac{3k}{6}\right)^2 \\ &= \left(\frac{9}{4} + \frac{2}{16} + \frac{3}{4}\right)k^2 \\ &= \frac{25}{8}k^2 \end{aligned}$$

so $k = \pm\frac{2\sqrt{2}}{5}$. The desired points are $\left(\frac{3\sqrt{2}}{5}, -\frac{\sqrt{2}}{10}, \frac{\sqrt{2}}{5}\right)$ and $\left(-\frac{3\sqrt{2}}{5}, \frac{\sqrt{2}}{10}, -\frac{\sqrt{2}}{5}\right)$. ♣

Exercises for 13.5

Exercise 13.5.1 *Find $D_{\boldsymbol{u}}f$ for $f = x^2 + xy + y^2$ in the direction of $\boldsymbol{u} = \langle 2, 1\rangle$ at the point $(1,1)$.*

Exercise 13.5.2 *Find $D_{\boldsymbol{u}}f$ for $f = \sin(xy)$ in the direction of $\boldsymbol{u} = \langle -1, 1\rangle$ at the point $(3,1)$.*

Exercise 13.5.3 *Find $D_{\boldsymbol{u}}f$ for $f = e^x\cos(y)$ in the direction 30 degrees from the positive x axis at the point $(1, \pi/4)$.*

Exercise 13.5.4 *The temperature of a thin plate in the x-y plane is $T = x^2 + y^2$. How fast does temperature change at the point $(1,5)$ moving in a direction 30 degrees from the positive x axis?*

Exercise 13.5.5 *Suppose the density of a thin plate at (x,y) is $1/\sqrt{x^2+y^2+1}$. Find the rate of change of the density at $(2,1)$ in a direction $\pi/3$ radians from the positive x axis.*

Exercise 13.5.6 *Suppose the electric potential at (x,y) is $\ln\sqrt{x^2+y^2}$. Find the rate of change of the potential at $(3,4)$ toward the origin and also in a direction at a right angle to the direction toward the origin.*

Exercise 13.5.7 *A plane perpendicular to the x-y plane contains the point $(2,1,8)$ on the paraboloid $z = x^2 + 4y^2$. The cross-section of the paraboloid created by this plane has slope 0 at this point. Find an equation of the plane.*

Exercise 13.5.8 *A plane perpendicular to the x-y plane contains the point $(3,2,2)$ on the paraboloid $36z = 4x^2 + 9y^2$. The cross-section of the paraboloid created by this plane has slope 0 at this point. Find an equation of the plane.*

Exercise 13.5.9 *Suppose the temperature at (x,y,z) is given by $T = xy + \sin(yz)$. In what direction should you go from the point $(1,1,1)$ to decrease the temperature as quickly as possible? What is the rate of change of temperature in this direction?*

Exercise 13.5.10 *Suppose the temperature at (x,y,z) is given by $T = xyz$. In what direction can you go from the point $(1,1,1)$ to maintain the same temperature?*

Exercise 13.5.11 *Find an equation for the plane tangent to $x^2 - 3y^2 + z^2 = 7$ at $(1,1,3)$.*

Exercise 13.5.12 *Find an equation for the plane tangent to $xyz = 6$ at $(1,2,3)$.*

Exercise 13.5.13 *Find an equation for the line normal to $x^2 + 2y^2 + 4z^2 = 26$ at $(2,-3,-1)$.*

Exercise 13.5.14 *Find an equation for the line normal to $x^2 + y^2 + 9z^2 = 56$ at $(4,2,-2)$.*

Exercise 13.5.15 *Find an equation for the line normal to $x^2 + 5y^2 - z^2 = 0$ at $(4,2,6)$.*

Exercise 13.5.16 *Find the directions in which the directional derivative of $f(x,y) = x^2 + \sin(xy)$ at the point $(1,0)$ has the value 1.*

Exercise 13.5.17 *Show that the curve $\mathbf{r}(t) = \langle \ln(t), t\ln(t), t \rangle$ is tangent to the surface $xz^2 - yz + \cos(xy) = 1$ at the point $(0,0,1)$.*

Exercise 13.5.18 *A bug is crawling on the surface of a hot plate, the temperature of which at the point x units to the right of the lower left corner and y units up from the lower left corner is given by $T(x,y) = 100 - x^2 - 3y^3$.*

(a) If the bug is at the point $(2,1)$, in what direction should it move to cool off the fastest? How fast will the temperature drop in this direction?

(b) If the bug is at the point $(1,3)$, in what direction should it move in order to maintain its temperature?

Exercise 13.5.19 *The elevation on a portion of a hill is given by $f(x,y) = 100 - 4x^2 - 2y$. From the location above $(2,1)$, in which direction will water run?*

Exercise 13.5.20 *Suppose that $g(x,y) = y - x^2$. Find the gradient at the point $(-1,3)$. Sketch the level curve to the graph of g when $g(x,y) = 2$, and plot both the tangent line and the gradient vector at the point $(-1,3)$. (Make your sketch large). What do you notice, geometrically?*

Exercise 13.5.21 *The gradient ∇f is a vector valued function of two variables. Prove the following gradient rules. Assume $f(x,y)$ and $g(x,y)$ are differentiable functions.*

(a) $\nabla(fg) = f\nabla(g) + g\nabla(f)$

(b) $\nabla(f/g) = (g\nabla f - f\nabla g)/g^2$

(c) $\nabla((f(x,y))^n) = nf(x,y)^{n-1}\nabla f$

13.6 Higher Order Derivatives

In single variable calculus we saw that the second derivative is often useful: in appropriate circumstances it measures acceleration; it can be used to identify maximum and minimum points; it tells us something about how sharply curved a graph is. Not surprisingly, second derivatives are also useful in the multi-variable case, but again not surprisingly, things are a bit more complicated.

It's easy to see where some complication is going to come from: with two variables there are four possible second derivatives. To take a "derivative," we must take a partial derivative with respect to x or y, and there are four ways to do it: x then x, x then y, y then x, y then y.

Example 13.22: Second Derivatives

Compute all four second derivatives of $f(x,y) = x^2y^2$.

Solution. Using an obvious notation, we get:

$$f_{xx} = 2y^2 \qquad f_{xy} = 4xy \qquad f_{yx} = 4xy \qquad f_{yy} = 2x^2.$$

♣

You will have noticed that two of these are the same, the "mixed partials" computed by taking partial derivatives with respect to both variables in the two possible orders. This is not an accident—as long as the function is reasonably nice, this will always be true.

Theorem 13.23: Clairaut's Theorem

If the mixed partial derivatives are continuous, they are equal.

Example 13.24: Mixed Partials

Compute the mixed partials of $f = xy/(x^2+y^2)$.

Solution. The mixed partial f_{xy} is found by first taking the partial derivative with respect to x:

$$f_x = \frac{y^3 - x^2y}{(x^2+y^2)^2},$$

then with respect to y:

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

We leave f_{yx} as an exercise. ♣

Exercises for 13.6

Exercise 13.6.1 *Let $f = xy/(x^2+y^2)$; compute f_{xx}, f_{yx}, and f_{yy}.*

Exercise 13.6.2 *Find all first and second partial derivatives of $x^3y^2+y^5$.*

Exercise 13.6.3 *Find all first and second partial derivatives of $4x^3+xy^2+10$.*

Exercise 13.6.4 *Find all first and second partial derivatives of $x\sin y$.*

Exercise 13.6.5 *Find all first and second partial derivatives of $\sin(3x)\cos(2y)$.*

Exercise 13.6.6 *Find all first and second partial derivatives of e^{x+y^2}.*

Exercise 13.6.7 *Find all first and second partial derivatives of $\ln\sqrt{x^3+y^4}$.*

Exercise 13.6.8 *Find all first and second partial derivatives of z with respect to x and y if $x^2+4y^2+16z^2-64=0$.*

Exercise 13.6.9 *Find all first and second partial derivatives of z with respect to x and y if $xy+yz+xz=1$.*

Exercise 13.6.10 *Let α and k be constants. Prove that the function $u(x,t) = e^{-\alpha^2k^2t}\sin(kx)$ is a solution to the heat equation $u_t = \alpha^2 u_{xx}$*

Exercise 13.6.11 *Let a be a constant. Prove that $u = \sin(x-at)+\ln(x+at)$ is a solution to the wave equation $u_{tt} = a^2u_{xx}$.*

Exercise 13.6.12 *How many third-order derivatives does a function of 2 variables have? How many of these are distinct?*

Exercise 13.6.13 *How many nth order derivatives does a function of 2 variables have? How many of these are distinct?*

13.7 Maxima and Minima

Suppose a surface given by $f(x,y)$ has a local maximum at (x_0,y_0,z_0); geometrically, this point on the surface looks like the top of a hill. If we look at the cross-section in the plane $y=y_0$, we will see a local maximum on the curve at (x_0,z_0), and we know from single-variable calculus that $\frac{\partial z}{\partial x}=0$ at this point. Likewise, in the plane $x=x_0$, $\frac{\partial z}{\partial y}=0$. So if there is a local maximum at (x_0,y_0,z_0), both partial derivatives at the point must be zero, and likewise for a local minimum. Thus, to find local maximum and minimum points, we need only consider those points at which both partial derivatives are 0. As in the single-variable case, it is possible for the derivatives to be 0 at a point that is neither a maximum or a minimum, so we need to test these points further.

You will recall that in the single variable case, we examined three methods to identify maximum and minimum points; the most useful is the second derivative test, though it does not always work. For functions of two variables there is also a second derivative test; again it is by far the most useful test, though it doesn't always work.

Theorem 13.25: Extrema Test for Multivariate Functions

Suppose that the second partial derivatives of $f(x,y)$ are continuous near (x_0,y_0), and $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$. We denote by D the **discriminant** *$D(x_0,y_0) = f_{xx}(x_0,y_0)f_{yy}(x_0,y_0) - f_{xy}(x_0,y_0)^2$. If $D > 0$ and $f_{xx}(x_0,y_0) < 0$ there is a local maximum at (x_0,y_0); if $D > 0$ and $f_{xx}(x_0,y_0) > 0$ there is a local minimum at (x_0,y_0); if $D < 0$ there is neither a maximum nor a minimum at (x_0,y_0); if $D = 0$, the test fails.*

Example 13.26: Extrema on an Elliptic Paraboloid

Verify that $f(x,y) = x^2 + y^2$ has a minimum at $(0,0)$.

Solution. First, we compute all the needed derivatives:

$$f_x = 2x \qquad f_y = 2y \qquad f_{xx} = 2 \qquad f_{yy} = 2 \qquad f_{xy} = 0.$$

The derivatives f_x and f_y are zero only at $(0,0)$. Applying the second derivative test there:

$$D(0,0) = f_{xx}(0,0)f_{yy}(0,0) - f_{xy}(0,0)^2 = 2 \cdot 2 - 0 = 4 > 0,$$

so there is a local minimum at $(0,0)$, and there are no other possibilities. ♣

Example 13.27: Extrema on a Hyperbolic Paraboloid

Find all local maxima and minima for $f(x,y) = x^2 - y^2$.

Solution. The derivatives:

$$f_x = 2x \qquad f_y = -2y \qquad f_{xx} = 2 \qquad f_{yy} = -2 \qquad f_{xy} = 0.$$

Again there is a single critical point, at $(0,0)$, and

$$D(0,0) = f_{xx}(0,0)f_{yy}(0,0) - f_{xy}(0,0)^2 = 2 \cdot -2 - 0 = -4 < 0,$$

so there is neither a maximum nor minimum there, and so there are no local maxima or minima. The surface is shown in Figure 13.9. ♣

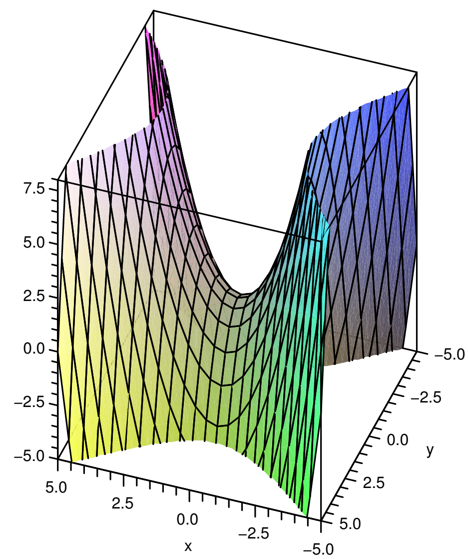


Figure 13.9: A saddle point, neither a maximum nor a minimum.

Example 13.28: Finding Extrema

Find all local maxima and minima for $f(x,y) = x^4 + y^4$.

Solution. The derivatives:

$$f_x = 4x^3 \qquad f_y = 4y^3 \qquad f_{xx} = 12x^2 \qquad f_{yy} = 12y^2 \qquad f_{xy} = 0.$$

Again there is a single critical point, at $(0,0)$, and

$$D(0,0) = f_{xx}(0,0)f_{yy}(0,0) - f_{xy}(0,0)^2 = 0 \cdot 0 - 0 = 0,$$

so we get no information. However, in this case it is easy to see that there is a minimum at $(0,0)$, because $f(0,0) = 0$ and at all other points $f(x,y) > 0$. ♣

Example 13.29: Finding Extrema

Find all local maxima and minima for $f(x,y) = x^3 + y^3$.

Solution. The derivatives:

$$f_x = 3x^2 \qquad f_y = 3y^2 \qquad f_{xx} = 6x^2 \qquad f_{yy} = 6y^2 \qquad f_{xy} = 0.$$

Again there is a single critical point, at $(0,0)$, and

$$D(0,0) = f_{xx}(0,0)f_{yy}(0,0) - f_{xy}(0,0)^2 = 0 \cdot 0 - 0 = 0,$$

so we get no information. In this case, a little thought shows there is neither a maximum nor a minimum at $(0,0)$: when x and y are both positive, $f(x,y) > 0$, and when x and y are both negative, $f(x,y) < 0$, and there are points of both kinds arbitrarily close to $(0,0)$. Alternately, if we look at the cross-section when $y = 0$, we get $f(x,0) = x^3$, which does not have either a maximum or minimum at $x = 0$. ♣

Example 13.30: Optimizing Dimensions of a Box

Suppose a box with no top is to hold a certain volume V. Find the dimensions for the box that result in the minimum surface area.

Solution. The area of the box is $A = 2hw + 2hl + lw$, and the volume is $V = lwh$, so we can write the area as a function of two variables,

$$A(l,w) = \frac{2V}{l} + \frac{2V}{w} + lw.$$

Then

$$A_l = -\frac{2V}{l^2} + w \quad \text{and} \quad A_w = -\frac{2V}{w^2} + l.$$

If we set these equal to zero and solve, we find $w = (2V)^{1/3}$ and $l = (2V)^{1/3}$, and the corresponding height is $h = V/(2V)^{2/3}$.

The second derivatives are

$$A_{ll} = \frac{4V}{l^3} \qquad A_{ww} = \frac{4V}{w^3} \qquad A_{lw} = 1,$$

so the discriminant is

$$D = \frac{4V}{l^3}\frac{4V}{w^3} - 1 = 4 - 1 = 3 > 0.$$

Since A_{ll} is 2, there is a local minimum at the critical point. Is this a global minimum? It is, but it is difficult to see this analytically; physically and graphically it is clear that there is a minimum, in which case it must be at the single critical point. ♣

Recall that when we did single variable global maximum and minimum problems, the easiest cases were those for which the variable could be limited to a finite closed interval, for then we simply had to check all critical values and the endpoints. The previous example is difficult because there is no finite boundary to the domain of the problem—both w and l can be in $(0,\infty)$. As in the single variable case, the problem is often simpler when there is a finite boundary.

Theorem 13.31: Multivariate Absolute Extrema

If $f(x,y)$ is continuous on a closed and bounded subset of $\mathbb{R}^2$, then it has both a maximum and minimum value.

As in the case of single variable functions, this means that the maximum and minimum values must occur at a critical point or on the boundary; in the two variable case, however, the boundary is a curve, not merely two endpoints.

Example 13.32: Optimizing Volume of a Box

The length of the diagonal of a box is to be 1 meter; find the maximum possible volume.

Solution. If the box is placed with one corner at the origin, and sides along the axes, the length of the diagonal is $\sqrt{x^2+y^2+z^2}$, and the volume is

$$V = xyz = xy\sqrt{1-x^2-y^2}.$$

Clearly, $x^2+y^2 \leq 1$, so the domain we are interested in is the quarter of the unit disk in the first quadrant. Computing derivatives:

$$V_x = \frac{y-2yx^2-y^3}{\sqrt{1-x^2-y^2}}$$
$$V_y = \frac{x-2xy^2-x^3}{\sqrt{1-x^2-y^2}}$$

If these are both 0, then $x=0$ or $y=0$, or $x=y=1/\sqrt{3}$. The boundary of the domain is composed of three curves: $x=0$ for $y \in [0,1]$; $y=0$ for $x \in [0,1]$; and $x^2+y^2=1$, where $x \geq 0$ and $y \geq 0$. In all three cases, the volume $xy\sqrt{1-x^2-y^2}$ is 0, so the maximum occurs at the only critical point $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$. See Figure 13.10. ♣

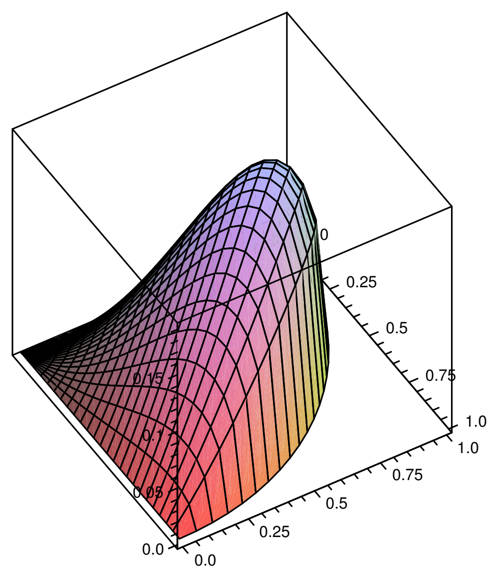


Figure 13.10: The volume of a box with fixed length diagonal.

Exercises for 13.7

Exercise 13.7.1 *Find all local maximum and minimum points of* $f = x^2 + 4y^2 - 2x + 8y - 1$.

Exercise 13.7.2 *Find all local maximum and minimum points of* $f = x^2 - y^2 + 6x - 10y + 2$.

Exercise 13.7.3 *Find all local maximum and minimum points of* $f = xy$.

Exercise 13.7.4 *Find all local maximum and minimum points of* $f = 9 + 4x - y - 2x^2 - 3y^2$.

Exercise 13.7.5 *Find all local maximum and minimum points of* $f = x^2 + 4xy + y^2 - 6y + 1$.

Exercise 13.7.6 *Find all local maximum and minimum points of* $f = x^2 - xy + 2y^2 - 5x + 6y - 9$.

Exercise 13.7.7 *Find the absolute maximum and minimum points of* $f = x^2 + 3y - 3xy$ *over the region bounded by* $y = x$, $y = 0$, *and* $x = 2$.

Exercise 13.7.8 *A six-sided rectangular box is to hold* $1/2$ *cubic meter; what shape should the box be to minimize surface area?*

Exercise 13.7.9 *The post office will accept packages whose combined length and girth is at most 130 inches. (Girth is the maximum distance around the package perpendicular to the length; for a rectangular box, the length is the largest of the three dimensions.) What is the largest volume that can be sent in a rectangular box?*

Exercise 13.7.10 *The bottom of a rectangular box costs twice as much per unit area as the sides and top. Find the shape for a given volume that will minimize cost.*

Exercise 13.7.11 *Using the methods of this section, find the shortest distance from the origin to the plane* $x + y + z = 10$.

Exercise 13.7.12 *Using the methods of this section, find the shortest distance from the point* (x_0, y_0, z_0) *to the plane* $ax + by + cz = d$. *You may assume that* $c \neq 0$; *use of Sage or similar software is recommended.*

Exercise 13.7.13 *A trough is to be formed by bending up two sides of a long metal rectangle so that the cross-section of the trough is an isosceles trapezoid. If the width of the metal sheet is 2 meters, how should it be bent to maximize the volume of the trough?*

Exercise 13.7.14 *Given the three points* $(1,4)$, $(5,2)$, *and* $(3,-2)$, $(x-1)^2 + (y-4)^2 + (x-5)^2 + (y-2)^2 + (x-3)^2 + (y+2)^2$ *is the sum of the squares of the distances from point* (x,y) *to the three points. Find* x *and* y *so that this quantity is minimized.*

Exercise 13.7.15 *Suppose that* $f(x,y) = x^2 + y^2 + kxy$. *Find and classify the critical points, and discuss how they change when* k *takes on different values.*

Exercise 13.7.16 *Find the shortest distance from the point* $(0,b)$ *to the parabola* $y=x^2$.

Exercise 13.7.17 *Find the shortest distance from the point* $(0,0,b)$ *to the paraboloid* $z=x^2+y^2$.

Exercise 13.7.18 *Consider the function* $f(x,y)=x^3-3x^2y+y^3$.

(a) *Show that* $(0,0)$ *is the only critical point of* f.

(b) *Show that the discriminant test is inconclusive for* f.

(c) *Determine the cross-sections of* f *obtained by setting* $y=kx$ *for various values of* k.

(d) *What kind of critical point is* $(0,0)$*?*

Exercise 13.7.19 *Find the volume of the largest rectangular box with edges parallel to the axes that can be inscribed in the ellipsoid* $2x^2+72y^2+18z^2=288$.

13.8 Lagrange Multipliers

Many applied max/min problems take the following form: we want to find an extreme value of a function, like $V=xyz$, subject to a constraint, like $1=\sqrt{x^2+y^2+z^2}$. Often this can be done, as we have, by explicitly combining the equations and then finding critical points. There is another approach that is often convenient, the method of **Lagrange multipliers**.

It is somewhat easier to understand two variable problems, so we begin with one as an example. Suppose the perimeter of a rectangle is to be 100 units. Find the rectangle with largest area. This is a fairly straightforward problem from single variable calculus. We write down the two equations: $A=xy$, $P=100=2x+2y$, solve the second of these for y (or x), substitute into the first, and end up with a one-variable maximization problem. Let's now think of it differently: the equation $A=xy$ defines a surface, and the equation $100=2x+2y$ defines a curve (a line, in this case) in the x-y plane. If we graph both of these in the three-dimensional coordinate system, we can phrase the problem like this: what is the highest point on the surface above the line? The solution we already understand effectively produces the equation of the cross-section of the surface above the line and then treats it as a single variable problem. Instead, imagine that we draw the level curves (the contour lines) for the surface in the x-y plane, along with the line.

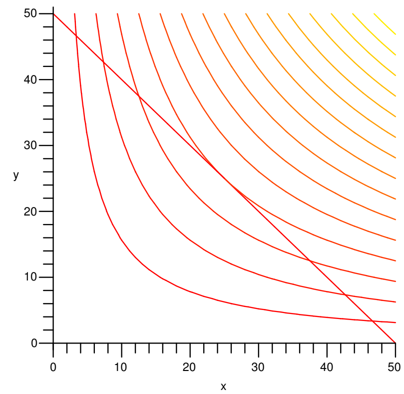


Figure 13.11: Constraint line with contour plot of the surface xy.

Imagine that the line represents a hiking trail and the contour lines are, as on a topographic map, the lines of constant altitude. How could you estimate, based on the graph, the high (or low) points on the path? As the path crosses contour lines, you know the path must be increasing or decreasing in elevation. At some point you will see the path just touch a contour line (tangent to it), and then begin to cross contours in the opposite order—that point of tangency must be a maximum or minimum point. If we can identify all such points, we can then check them to see which gives the maximum and which the minimum value. As usual, we also need to check boundary points; in this problem, we know that x and y are positive, so we are interested in just the portion of the line in the first quadrant, as shown. The endpoints of the path, the two points on the axes, are not points of tangency, but they are the two places that the function xy is a minimum in the first quadrant.

How can we actually make use of this? At the points of tangency that we seek, the constraint curve (in this case the line) and the level curve have the same slope—their tangent lines are parallel. This also means that the constraint curve is perpendicular to the gradient vector of the function; going a bit further, if we can express the constraint curve itself as a level curve, then we seek the points at which the two level curves have parallel gradients. The curve $100 = 2x + 2y$ can be thought of as a level curve of the function $2x + 2y$; Figure 13.12 shows both sets of level curves on a single graph. We are interested in those points where two level curves are tangent—but there are many such points, in fact an infinite number, as we've only shown a few of the level curves. All along the line $y = x$ are points at which two level curves are tangent. While this might seem to be a show-stopper, it is not.

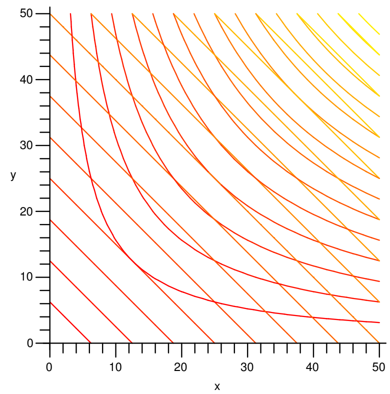


Figure 13.12: Contour plots for $2x+2y$ and xy.

The gradient of $2x+2y$ is $\langle 2,2\rangle$, and the gradient of xy is $\langle y,x\rangle$. They are parallel when $\langle 2,2\rangle = \lambda\langle y,x\rangle$, that is, when $2=\lambda y$ and $2=\lambda x$. We have two equations in three unknowns, which typically results in many solutions (as we expected). A third equation will reduce the number of solutions; the third equation is the original constraint, $100=2x+2y$. So we have the following system to solve:

$$2=\lambda y \qquad 2=\lambda x \qquad 100=2x+2y.$$

In the first two equations, λ can't be 0, so we may divide by it to get $x=y=2/\lambda$. Substituting into the third equation we get

$$\begin{aligned} 2\frac{2}{\lambda}+2\frac{2}{\lambda} &= 100 \\ \frac{8}{100} &= \lambda \end{aligned}$$

so $x=y=25$. Note that we are not really interested in the value of λ—it is a clever tool, the Lagrange multiplier, introduced to solve the problem. In many cases, as here, it is easier to find λ than to find everything else without using λ.

The same method works for functions of three variables, except of course everything is one dimension higher: the function to be optimized is a function of three variables and the constraint represents a surface—for example, the function may represent temperature, and we may be interested in the maximum temperature on some surface, like a sphere. The points we seek are those at which the constraint surface is tangent to a level surface of the function. Once again, we consider the constraint surface to be a level surface of some function, and we look for points at which the two gradients are parallel, giving us three equations in four unknowns. The constraint provides a fourth equation.

Example 13.33: Optimization with Constraints

Maximize the function xyz *given the constraint* $1=\sqrt{x^2+y^2+z^2}$.

Solution. The constraint is $1 = \sqrt{x^2+y^2+z^2}$, which is the same as $1 = x^2+y^2+z^2$. The function to maximize is xyz. The two gradient vectors are $\langle 2x, 2y, 2z\rangle$ and $\langle yz, xz, xy\rangle$, so the equations to be solved are

$$\begin{aligned} yz &= 2x\lambda \\ xz &= 2y\lambda \\ xy &= 2z\lambda \\ 1 &= x^2+y^2+z^2 \end{aligned}$$

If $\lambda = 0$ then at least two of x, y, z must be 0, giving a volume of 0, which will not be the maximum. If we multiply the first two equations by x and y respectively, we get

$$\begin{aligned} xyz &= 2x^2\lambda \\ xyz &= 2y^2\lambda \end{aligned}$$

so $2x^2\lambda = 2y^2\lambda$ or $x^2 = y^2$; in the same way we can show $x^2 = z^2$. Hence the fourth equation becomes $1 = x^2+x^2+x^2$ or $x = 1/\sqrt{3}$, and so $x = y = z = 1/\sqrt{3}$ gives the maximum volume. This is of course the same answer we obtained previously. ♣

Another possibility is that we have a function of three variables, and we want to find a maximum or minimum value not on a surface but on a curve; often the curve is the intersection of two surfaces, so that we really have two constraint equations, say $g(x,y,z) = c_1$ and $h(x,y,z) = c_2$. It turns out that at points on the intersection of the surfaces where f has a maximum or minimum value,

$$\nabla f = \lambda \nabla g + \mu \nabla h.$$

As before, this gives us three equations, one for each component of the vectors, but now in five unknowns, x, y, z, λ, and μ. Since there are two constraint functions, we have a total of five equations in five unknowns, and so can usually find the solutions we need.

Example 13.34: Intersection of a Plane with a Cylinder

The plane $x+y-z = 1$ intersects the cylinder $x^2+y^2 = 1$ in an ellipse. Find the points on the ellipse closest to and farthest from the origin.

Solution. We want the extreme values of $f = \sqrt{x^2+y^2+z^2}$ subject to the constraints $g = x^2+y^2 = 1$ and $h = x+y-z = 1$. To simplify the algebra, we may use instead $f = x^2+y^2+z^2$, since this has a maximum or minimum value at exactly the points at which $\sqrt{x^2+y^2+z^2}$ does. The gradients are

$$\nabla f = \langle 2x, 2y, 2z\rangle \qquad \nabla g = \langle 2x, 2y, 0\rangle \qquad \nabla h = \langle 1, 1, -1\rangle,$$

so the equations we need to solve are

$$\begin{aligned} 2x &= \lambda 2x + \mu \\ 2y &= \lambda 2y + \mu \end{aligned}$$

$$\begin{aligned} 2z &= 0 - \mu \\ 1 &= x^2 + y^2 \\ 1 &= x + y - z. \end{aligned}$$

Subtracting the first two we get $2y - 2x = \lambda(2y - 2x)$, so either $\lambda = 1$ or $x = y$. If $\lambda = 1$ then $\mu = 0$, so $z = 0$ and the last two equations are

$$1 = x^2 + y^2 \qquad \text{and} \qquad 1 = x + y.$$

Solving these gives $x = 1$, $y = 0$, or $x = 0$, $y = 1$, so the points of interest are $(1,0,0)$ and $(0,1,0)$, which are both distance 1 from the origin. If $x = y$, the fourth equation is $2x^2 = 1$, giving $x = y = \pm 1/\sqrt{2}$, and from the fifth equation we get $z = -1 \pm \sqrt{2}$. The distance from the origin to $(1/\sqrt{2}, 1/\sqrt{2}, -1+\sqrt{2})$ is $\sqrt{4 - 2\sqrt{2}} \approx 1.08$ and the distance from the origin to $(-1/\sqrt{2}, -1/\sqrt{2}, -1-\sqrt{2})$ is $\sqrt{4+2\sqrt{2}} \approx 2.6$. Thus, the points $(1,0,0)$ and $(0,1,0)$ are closest to the origin and $(-1/\sqrt{2}, -1/\sqrt{2}, -1-\sqrt{2})$ is farthest from the origin. ♣

Exercises for 13.8

Exercise 13.8.1 *A six-sided rectangular box is to hold $1/2$ cubic meter; what shape should the box be to minimize surface area?*

Exercise 13.8.2 *The post office will accept packages whose combined length and girth are at most 130 inches (girth is the maximum distance around the package perpendicular to the length). What is the largest volume that can be sent in a rectangular box?*

Exercise 13.8.3 *The bottom of a rectangular box costs twice as much per unit area as the sides and top. Find the shape for a given volume that will minimize cost.*

Exercise 13.8.4 *Using Lagrange multipliers, find the shortest distance from the point (x_0, y_0, z_0) to the plane $ax + by + cz = d$.*

Exercise 13.8.5 *Find all points on the surface $xy - z^2 + 1 = 0$ that are closest to the origin.*

Exercise 13.8.6 *The material for the bottom of an aquarium costs half as much as the high strength glass for the four sides. Find the shape of the cheapest aquarium that holds a given volume V.*

Exercise 13.8.7 *The plane $x - y + z = 2$ intersects the cylinder $x^2 + y^2 = 4$ in an ellipse. Find the points on the ellipse closest to and farthest from the origin.*

Exercise 13.8.8 *Find three positive numbers whose sum is 48 and whose product is as large as possible.*

Exercise 13.8.9 *Find all points on the plane* $x+y+z=5$ *in the first octant at which* $f(x,y,z)=xy^2z^2$ *has a maximum value.*

Exercise 13.8.10 *Find the points on the surface* $x^2-yz=5$ *that are closest to the origin.*

Exercise 13.8.11 *A manufacturer makes two models of an item, standard and deluxe. It costs \$40 to manufacture the standard model and \$60 for the deluxe. A market research firm estimates that if the standard model is priced at x dollars and the deluxe at y dollars, then the manufacturer will sell* $500(y-x)$ *of the standard items and* $45,000+500(x-2y)$ *of the deluxe each year. How should the items be priced to maximize profit?*

Exercise 13.8.12 *A length of sheet metal is to be made into a water trough by bending up two sides as shown in Figure 13.13. Find x and* ϕ *so that the trapezoid–shaped cross section has maximum area, when the width of the metal sheet is 27 inches (that is,* $2x+y=27$*).*

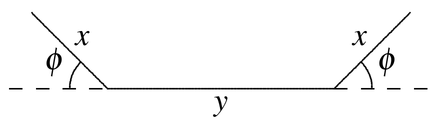


Figure 13.13: Cross-section of a trough.

Exercise 13.8.13 *Find the maximum and minimum values of* $f(x,y,z)=6x+3y+2z$ *subject to the constraint* $g(x,y,z)=4x^2+2y^2+z^2-70=0$.

Exercise 13.8.14 *Find the maximum and minimum values of* $f(x,y)=e^{xy}$ *subject to the constraint* $g(x,y)=x^3+y^3-16=0$.

Exercise 13.8.15 *Find the maximum and minimum values of* $f(x,y)=xy+\sqrt{9-x^2-y^2}$ *when* $x^2+y^2\leq 9$.

Exercise 13.8.16 *Find three real numbers whose sum is 9 and the sum of whose squares is a small as possible.*

Exercise 13.8.17 *Find the dimensions of the closed rectangular box with maximum volume that can be inscribed in the unit sphere.*

14. Multiple Integration

14.1 Volume and Average Height

Consider a surface $f(x,y)$; you might temporarily think of this as representing physical topography—a hilly landscape, perhaps. What is the average height of the surface (or average altitude of the landscape) over some region?

As with most such problems, we start by thinking about how we might approximate the answer. Suppose the region is a rectangle, $[a,b]\times[c,d]$. We can divide the rectangle into a grid, m subdivisions in one direction and n in the other, as indicated in Figure 14.1. We pick x values x_0, $x_1,\ldots$, x_{m-1} in each subdivision in the x direction, and similarly in the y direction. At each of the points (x_i,y_j) in one of the smaller rectangles in the grid, we compute the height of the surface: $f(x_i,y_j)$. Now the average of these heights should be (depending on the fineness of the grid) close to the average height of the surface:

$$\frac{f(x_0,y_0)+f(x_1,y_0)+\cdots+f(x_0,y_1)+f(x_1,y_1)+\cdots+f(x_{m-1},y_{n-1})}{mn}$$

As both m and n go to infinity, we expect this approximation to converge to a fixed value, the actual average height of the surface. For reasonably nice functions this does indeed happen.

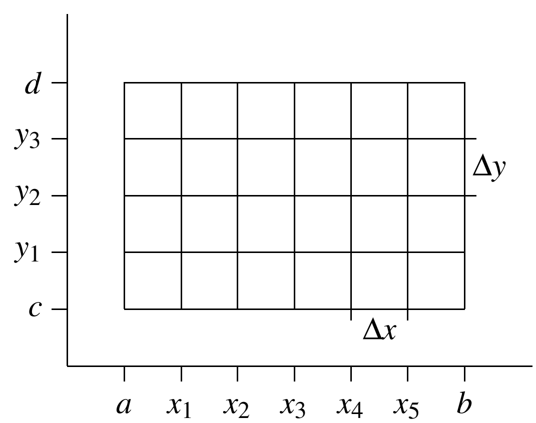


Figure 14.1: A rectangular subdivision of $[a,b]\times[c,d]$**.**

Using sigma notation, we can rewrite the approximation:

$$\begin{aligned}\frac{1}{mn}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} f(x_j,y_i) &= \frac{1}{(b-a)(d-c)}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} f(x_j,y_i)\frac{b-a}{m}\frac{d-c}{n}\\ &= \frac{1}{(b-a)(d-c)}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} f(x_j,y_i)\Delta x\Delta y.\end{aligned}$$

The two parts of this product have useful meaning: $(b-a)(d-c)$ is of course the area of the rectangle, and the double sum adds up mn terms of the form $f(x_j,y_i)\Delta x\Delta y$, which is the height of the surface at a point multiplied by the area of one of the small rectangles into which we have divided the large rectangle. In short, each term $f(x_j,y_i)\Delta x\Delta y$ is the volume of a tall, thin, rectangular box, and is approximately the volume under the surface and above one of the small rectangles; see Figure 14.2. When we add all of these up, we get an approximation to the volume under the surface and above the rectangle $R=[a,b]\times[c,d]$. When we take the limit as m and n go to infinity, the double sum becomes the actual volume under the surface, which we divide by $(b-a)(d-c)$ to get the average height.

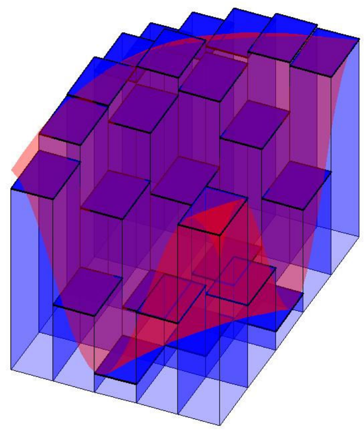


Figure 14.2: Approximating the volume under a surface.

Double sums like this come up in many applications, so in a way it is the most important part of this example; dividing by $(b-a)(d-c)$ is a simple extra step that allows the computation of an average. As we did in the single variable case, we introduce a special notation for the limit of such a double sum:

$$\lim_{m,n\to\infty}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} f(x_j,y_i)\Delta x\Delta y=\iint_R f(x,y)\,dx\,dy=\iint_R f(x,y)\,dA,$$

the **double integral** of f over the region R. The notation dA indicates a small bit of area, without specifying any particular order for the variables x and y; it is shorter and more "generic" than writing $dx\,dy$. The average height of the surface in this notation is

$$\frac{1}{(b-a)(d-c)}\iint_R f(x,y)\,dA.$$

The next question, of course, is: How do we compute these double integrals? You might think that we will need some two-dimensional version of the Fundamental Theorem of Calculus, but as it turns out we can get away with just the single variable version, applied twice.

Going back to the double sum, we can rewrite it to emphasize a particular order in which we want to add the terms:

$$\sum_{i=0}^{n-1}\left(\sum_{j=0}^{m-1} f(x_j,y_i)\Delta x\right)\Delta y.$$

In the sum in parentheses, only the value of x_j is changing; y_i is temporarily constant. As m goes to infinity, this sum has the right form to turn into an integral:

$$\lim_{m\to\infty}\sum_{j=0}^{m-1} f(x_j,y_i)\Delta x = \int_a^b f(x,y_i)\,dx.$$

So after we take the limit as m goes to infinity, the sum is

$$\sum_{i=0}^{n-1}\left(\int_a^b f(x,y_i)\,dx\right)\Delta y.$$

Of course, for different values of y_i this integral has different values; in other words, it is really a function applied to y_i:

$$G(y) = \int_a^b f(x,y)\,dx.$$

If we substitute back into the sum we get

$$\sum_{i=0}^{n-1} G(y_i)\Delta y.$$

This sum has a nice interpretation. The value $G(y_i)$ is the area of a cross section of the region under the surface $f(x,y)$, namely, when $y = y_i$. The quantity $G(y_i)\Delta y$ can be interpreted as the volume of a solid with face area $G(y_i)$ and thickness Δy. Think of the surface $f(x,y)$ as the top of a loaf of sliced bread. Each slice has a cross-sectional area and a thickness; $G(y_i)\Delta y$ corresponds to the volume of a single slice of bread. Adding these up approximates the total volume of the loaf. (This is very similar to the technique we used to compute volumes in Section 8.3, except that there we need the cross-sections to be in some way "the same".) Figure 14.3 shows this "sliced loaf" approximation using the same surface as shown in Figure 14.2. Nicely enough, this sum looks just like the sort of sum that turns into an integral, namely,

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=0}^{n-1} G(y_i)\Delta y &= \int_c^d G(y)\,dy\\ &= \int_c^d\int_a^b f(x,y)\,dx\,dy.\end{aligned}$$

Let's be clear about what this means: we first will compute the inner integral, temporarily treating y as a constant. We will do this by finding an anti-derivative with respect to x, then substituting $x = a$ and $x = b$ and subtracting, as usual. The result will be an expression with no x variable but some occurrences of y. Then the outer integral will be an ordinary one-variable problem, with y as the variable.

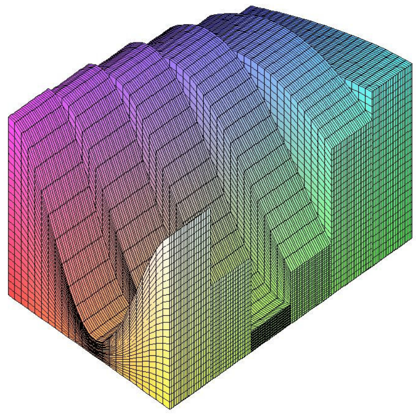


Figure 14.3: Approximating the volume under a surface with slices.

Example 14.1: Volume Under Surface

Figure 14.2 shows the function $\sin(xy) + 6/5$ *on* $[0.5, 3.5] \times [0.5, 2.5]$. *Find the volume under this surface.*

Solution. The volume under this surface is

$$\int_{0.5}^{2.5} \int_{0.5}^{3.5} \sin(xy) + \frac{6}{5}\, dx\, dy.$$

The inner integral is

$$\int_{0.5}^{3.5} \sin(xy) + \frac{6}{5}\, dx = \frac{-\cos(xy)}{y} + \frac{6x}{5}\bigg|_{0.5}^{3.5} = \frac{-\cos(3.5y)}{y} + \frac{\cos(0.5y)}{y} + \frac{18}{5}.$$

Unfortunately, this gives a function for which we can't find a simple anti-derivative. To complete the problem we could use Sage or similar software to approximate the integral. Doing this gives a volume of approximately 8.84, so the average height is approximately $8.84/6 \approx 1.47$. ♣

Because addition and multiplication are commutative and associative, we can rewrite the original double sum:

$$\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} f(x_j, y_i)\Delta x \Delta y = \sum_{j=0}^{m-1} \sum_{i=0}^{n-1} f(x_j, y_i)\Delta y \Delta x.$$

Now if we repeat the development above, the inner sum turns into an integral:

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(x_j, y_i)\Delta y = \int_c^d f(x_j, y)\, dy,$$

and then the outer sum turns into an integral:

$$\lim_{m\to\infty} \sum_{j=0}^{m-1} \left(\int_c^d f(x_j, y)\, dy \right) \Delta x = \int_a^b \int_c^d f(x, y)\, dy\, dx.$$

In other words, we can compute the integrals in either order, first with respect to x then y, or vice versa. Thinking of the loaf of bread, this corresponds to slicing the loaf in a direction perpendicular to the first.

We haven't really proved that the value of a double integral is equal to the value of the corresponding two single integrals in either order of integration, but provided the function is reasonably nice, this is true; the result is called **Fubini's Theorem**.

Example 14.2: Compute Volume in Two Ways

We compute $\iint_R 1+(x-1)^2+4y^2\,dA$, *where* $R=[0,3]\times[0,2]$, *in two ways.*

Solution. First,

$$\begin{aligned}
\int_0^3\int_0^2 1+(x-1)^2+4y^2\,dy\,dx &= \int_0^3 y+(x-1)^2y+\frac{4}{3}y^3\Big|_0^2\,dx \\
&= \int_0^3 2+2(x-1)^2+\frac{32}{3}\,dx \\
&= 2x+\frac{2}{3}(x-1)^3+\frac{32}{3}x\Big|_0^3 \\
&= 6+\frac{2}{3}\cdot 8+\frac{32}{3}\cdot 3-(0-1\cdot\frac{2}{3}+0) \\
&= 44.
\end{aligned}$$

In the other order:

$$\begin{aligned}
\int_0^2\int_0^3 1+(x-1)^2+4y^2\,dx\,dy &= \int_0^2 x+\frac{(x-1)^3}{3}+4y^2x\Big|_0^3\,dy \\
&= \int_0^2 3+\frac{8}{3}+12y^2+\frac{1}{3}\,dy \\
&= 3y+\frac{8}{3}y+4y^3+\frac{1}{3}y\Big|_0^2 \\
&= 6+\frac{16}{3}+32+\frac{2}{3} \\
&= 44.
\end{aligned}$$

♣

In this example there is no particular reason to favor one direction over the other; in some cases, one direction might be much easier than the other, so it's usually worth considering the two different possibilities.

Frequently we will be interested in a region that is not simply a rectangle. Let's compute the volume under the surface $x+2y^2$ above the region described by $0\le x\le 1$ and $0\le y\le x^2$, shown in Figure 14.4.

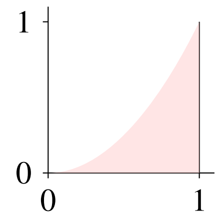


Figure 14.4: A parabolic region of integration.

In principle there is nothing more difficult about this problem. If we imagine the three-dimensional region under the surface and above the parabolic region as an oddly shaped loaf of bread, we can still slice it up, approximate the volume of each slice, and add these volumes up. For example, if we slice perpendicular to the x axis at x_i, the thickness of a slice will be Δx and the area of the slice will be

$$\int_0^{x_i^2} x_i + 2y^2\,dy.$$

When we add these up and take the limit as Δx goes to 0, we get the double integral

$$\begin{aligned}
\int_0^1 \int_0^{x^2} x + 2y^2\,dy\,dx &= \int_0^1 xy + \frac{2}{3}y^3\Big|_0^{x^2}\,dx \\
&= \int_0^1 x^3 + \frac{2}{3}x^6\,dx \\
&= \frac{x^4}{4} + \frac{2}{21}x^7\Big|_0^1 \\
&= \frac{1}{4} + \frac{2}{21} = \frac{29}{84}.
\end{aligned}$$

We could just as well slice the solid perpendicular to the y axis, in which case we get

$$\begin{aligned}
\int_0^1 \int_{\sqrt{y}}^1 x + 2y^2\,dx\,dy &= \int_0^1 \frac{x^2}{2} + 2y^2x\Big|_{\sqrt{y}}^1\,dy \\
&= \int_0^1 \frac{1}{2} + 2y^2 - \frac{y}{2} - 2y^2\sqrt{y}\,dy \\
&= \frac{y}{2} + \frac{2}{3}y^3 - \frac{y^2}{4} - \frac{4}{7}y^{7/2}\Big|_0^1 \\
&= \frac{1}{2} + \frac{2}{3} - \frac{1}{4} - \frac{4}{7} = \frac{29}{84}.
\end{aligned}$$

What is the average height of the surface over this region? As before, it is the volume divided by the area of the base, but now we need to use integration to compute the area of the base, since it is not a simple rectangle. The area is

$$\int_0^1 x^2\,dx = \frac{1}{3},$$

so the average height is $29/28$.

Example 14.3: Volume of Region

Find the volume under the surface $z = \sqrt{1-x^2}$ and above the triangle formed by $y = x$, $x = 1$, and the x-axis.

Solution. Let's consider the two possible ways to set this up:

$$\int_0^1 \int_0^x \sqrt{1-x^2}\,dy\,dx \qquad \text{or} \qquad \int_0^1 \int_y^1 \sqrt{1-x^2}\,dx\,dy.$$

Which appears easier? In the first, the first (inner) integral is easy, because we need an anti-derivative with respect to y, and the entire integrand $\sqrt{1-x^2}$ is constant with respect to y. Of course, the second integral may be more difficult. In the second, the first integral is mildly unpleasant—a trig substitution. So let's try the first one, since the first step is easy, and see where that leaves us.

$$\int_0^1 \int_0^x \sqrt{1-x^2}\,dy\,dx = \int_0^1 y\sqrt{1-x^2}\Big|_0^x\,dx = \int_0^1 x\sqrt{1-x^2}\,dx.$$

This is quite easy, since the substitution $u = 1-x^2$ works:

$$\int x\sqrt{1-x^2}\,dx = -\frac{1}{2}\int \sqrt{u}\,du = \frac{1}{3}u^{3/2} = -\frac{1}{3}(1-x^2)^{3/2}.$$

Then

$$\int_0^1 x\sqrt{1-x^2}\,dx = -\frac{1}{3}(1-x^2)^{3/2}\Big|_0^1 = \frac{1}{3}.$$

This is a good example of how the order of integration can affect the complexity of the problem. In this case it is possible to do the other order, but it is a bit messier. In some cases one order may lead to a very difficult or impossible integral; it's usually worth considering both possibilities before going very far. ♣

Exercises for 14.1

Exercise 14.1.1 *Compute* $\displaystyle\int_0^2 \int_0^4 1 + x\,dy\,dx.$

Exercise 14.1.2 *Compute* $\displaystyle\int_{-1}^1 \int_0^2 x + y\,dy\,dx.$

Exercise 14.1.3 *Compute* $\displaystyle\int_1^2 \int_0^y xy\,dx\,dy.$

Exercise 14.1.4 *Compute* $\displaystyle\int_0^1 \int_{y^2/2}^{\sqrt{y}} dx\,dy.$

Exercise 14.1.5 *Compute* $\displaystyle\int_1^2 \int_1^x \frac{x^2}{y^2}\,dy\,dx.$

Exercise 14.1.6 *Compute* $\displaystyle\int_0^1 \int_0^{x^2} \frac{y}{e^x}\,dy\,dx.$

Exercise 14.1.7 *Compute* $\displaystyle\int_0^{\sqrt{\pi/2}} \int_0^{x^2} x\cos y\,dy\,dx.$

Exercise 14.1.8 *Compute* $\displaystyle\int_0^{\pi/2} \int_0^{\cos\theta} r^2(\cos\theta - r)\,dr\,d\theta.$

Exercise 14.1.9 *Compute:* $\displaystyle\int_0^1 \int_{\sqrt{y}}^1 \sqrt{x^3+1}\,dx\,dy.$

Exercise 14.1.10 *Compute:* $\displaystyle\int_0^1 \int_{y^2}^1 y\sin(x^2)\,dx\,dy.$

Exercise 14.1.11 *Compute:* $\displaystyle\int_0^1 \int_{x^2}^1 x\sqrt{1+y^2}\,dy\,dx$

Exercise 14.1.12 *Compute:* $\displaystyle\int_0^1 \int_0^y \frac{2}{\sqrt{1-x^2}}\,dx\,dy$

Exercise 14.1.13 *Compute:* $\displaystyle\int_0^1 \int_{3y}^3 e^{x^2}\,dx\,dy$

Exercise 14.1.14 *Compute* $\displaystyle\int_{-1}^1 \int_0^{1-x^2} x^2 - \sqrt{y}\,dy\,dx.$

Exercise 14.1.15 *Compute* $\displaystyle\int_0^{\sqrt{2}/2} \int_{-\sqrt{1-2x^2}}^{\sqrt{1-2x^2}} x\,dy\,dx.$

Exercise 14.1.16 *Evaluate* $\displaystyle\iint x^2\,dA$ *over the region in the first quadrant bounded by the hyperbola* $xy = 16$ *and the lines* $y = x$, $y = 0$, *and* $x = 8$.

Exercise 14.1.17 *Find the volume below* $z = 1 - y$ *above the region* $-1 \le x \le 1$, $0 \le y \le 1 - x^2$.

Exercise 14.1.18 *Find the volume bounded by* $z = x^2 + y^2$ *and* $z = 4$.

Exercise 14.1.19 *Find the volume in the first octant bounded by* $y^2 = 4 - x$ *and* $y = 2z$.

Exercise 14.1.20 *Find the volume in the first octant bounded by* $y^2 = 4x$, $2x + y = 4$, $z = y$, *and* $y = 0$.

Exercise 14.1.21 *Find the volume in the first octant bounded by* $x + y + z = 9$, $2x + 3y = 18$, *and* $x + 3y = 9$.

Exercise 14.1.22 *Find the volume in the first octant bounded by $x^2+y^2=a^2$ and $z=x+y$.*

Exercise 14.1.23 *Find the volume bounded by $4x^2+y^2=4z$ and $z=2$.*

Exercise 14.1.24 *Find the volume bounded by $z=x^2+y^2$ and $z=y$.*

Exercise 14.1.25 *Find the volume under the surface $z=xy$ above the triangle with vertices $(1,1,0)$, $(4,1,0)$, $(1,2,0)$.*

Exercise 14.1.26 *Find the volume enclosed by $y=x^2$, $y=4$, $z=x^2$, $z=0$.*

Exercise 14.1.27 *A swimming pool is circular with a 40 meter diameter. The depth is constant along east-west lines and increases linearly from 2 meters at the south end to 7 meters at the north end. Find the volume of the pool.*

Exercise 14.1.28 *Find the average value of $f(x,y)=e^y\sqrt{x+e^y}$ on the rectangle with vertices $(0,0)$, $(4,0)$, $(4,1)$ and $(0,1)$.*

Exercise 14.1.29 *Figure 14.5 shows a temperature map of Colorado. Use the data to estimate the average temperature in the state using 4, 16 and 25 subdivisions. Give both an upper and lower estimate. Why do we like Colorado for this problem? What other state(s) might we like?*

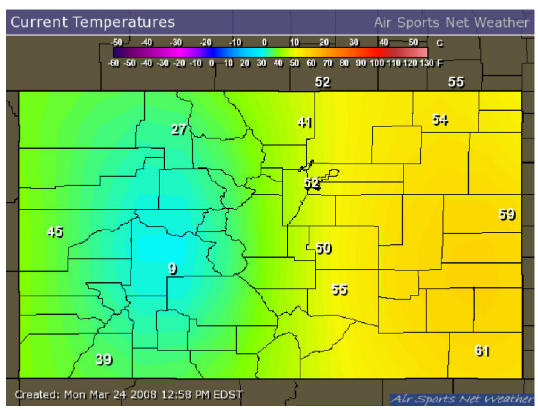


Figure 14.5: Colorado temperatures.

Exercise 14.1.30 *Three cylinders of radius 1 intersect at right angles at the origin, as shown in Figure 14.6. Find the volume contained inside all three cylinders.*

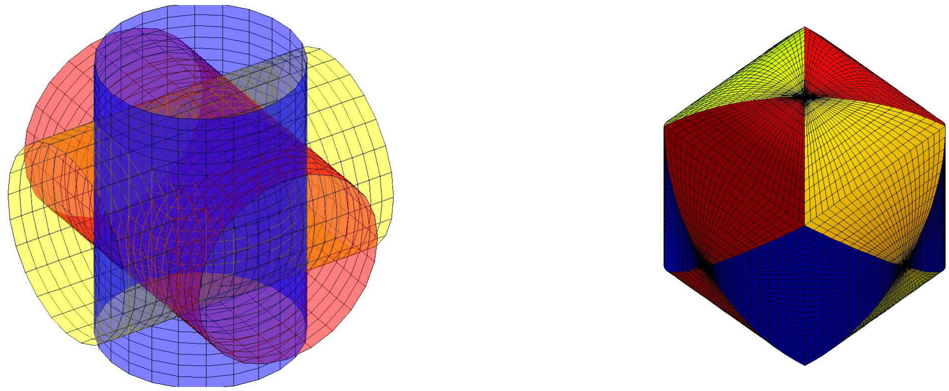


Figure 14.6: Intersection of three cylinders.

Exercise 14.1.31 *Prove that if $f(x,y)$ is integrable and if $g(x,y) = \int_a^x \int_b^y f(s,t)\,dt\,ds$ then $g_{xy} = g_{yx} = f(x,y)$.*

Exercise 14.1.32 *Reverse the order of integration on each of the following integrals*

(a) $\int_0^9 \int_0^{\sqrt{9-y}} f(x,y)\,dx\,dy$

(b) $\int_1^2 \int_0^{\ln x} f(x,y)\,dy\,dx$

(c) $\int_0^1 \int_{\arcsin y}^{\pi/2} f(x,y)\,dx\,dy$

(d) $\int_0^1 \int_{4x}^4 f(x,y)\,dy\,dx$

(e) $\int_0^3 \int_0^{\sqrt{9-y^2}} f(x,y)\,dx\,dy$

Exercise 14.1.33 *What are the parallels between Fubini's Theorem and Clairaut's Theorem?*

14.2 Double Integrals in Polar Coordinates

Suppose we have a surface given in polar coordinates as $z = f(r,\theta)$ and we wish to find the integral over some region. We could attempt to translate into rectangular coordinates and do the integration there, but it is often easier to stay in polar coordinates.

How might we approximate the volume under such a surface in a way that uses polar coordinates directly? The basic idea is the same as before: we divide the region into many small regions, multiply the area of each small region by the height of the surface somewhere in that little region, and add them up. What changes is the shape of the small regions; in order to have a nice representation in terms of r and θ, we use small pieces of ring-shaped areas, as shown in Figure 14.7. Each small region is roughly rectangular, except that two sides are segments of a circle and the other two sides are not quite parallel. Near a point (r,θ), the length of either circular arc is about $r\Delta\theta$ and the length of each straight side is simply Δr. When Δr and $\Delta\theta$ are very small, the region is nearly a rectangle with area $r\Delta r\Delta\theta$, and the volume under the surface is approximately

$$\sum\sum f(r_i,\theta_j)r_i\Delta r\Delta\theta.$$

In the limit, this turns into a double integral

$$\int_{\theta_0}^{\theta_1}\int_{r_0}^{r_1} f(r,\theta)r\,dr\,d\theta.$$

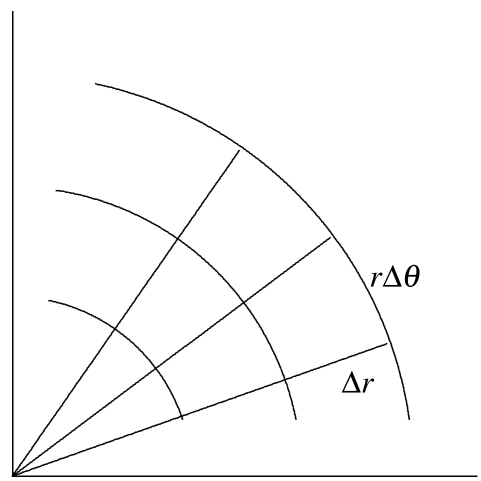


Figure 14.7: A polar coordinates "grid".

Example 14.4: Volume of One-Eighth of a Sphere

Find the volume under $z=\sqrt{4-r^2}$ above the quarter circle bounded by the two axes and the circle $x^2+y^2=4$ in the first quadrant.

Solution. In terms of r and θ, this region is described by the restrictions $0\le r\le 2$ and $0\le\theta\le\pi/2$, so we have

$$\begin{aligned}\int_0^{\pi/2}\int_0^2\sqrt{4-r^2}\,r\,dr\,d\theta &= \int_0^{\pi/2} -\frac{1}{3}(4-r^2)^{3/2}\Big|_0^2\,d\theta\\ &= \int_0^{\pi/2}\frac{8}{3}\,d\theta\\ &= \frac{4\pi}{3}.\end{aligned}$$

The surface is a portion of the sphere of radius 2 centered at the origin, in fact exactly one-eighth of the sphere. We know the formula for volume of a sphere is $(4/3)\pi r^3$, so the volume we have computed is $(1/8)(4/3)\pi 2^3 = (4/3)\pi$, in agreement with our answer. ♣

This example is much like a simple one in rectangular coordinates: the region of interest may be described exactly by a constant range for each of the variables. As with rectangular coordinates, we can adapt the method to deal with more complicated regions.

Example 14.5: Integration in Polar Coordinates

Find the volume under $z = \sqrt{4-r^2}$ above the region enclosed by the curve $r = 2\cos\theta$, $-\pi/2 \le \theta \le \pi/2$; see Figure 14.8.

Solution. The region is described in polar coordinates by the inequalities $-\pi/2 \le \theta \le \pi/2$ and $0 \le r \le 2\cos\theta$, so the double integral is

$$\int_{-\pi/2}^{\pi/2}\int_0^{2\cos\theta} \sqrt{4-r^2}\, r\,dr\,d\theta = 2\int_0^{\pi/2}\int_0^{2\cos\theta} \sqrt{4-r^2}\, r\,dr\,d\theta.$$

We can rewrite the integral as shown because of the symmetry of the volume; this avoids a complication during the evaluation. Proceeding:

$$\begin{aligned}
2\int_0^{\pi/2}\int_0^{2\cos\theta} \sqrt{4-r^2}\, r\,dr\,d\theta &= 2\int_0^{\pi/2} -\frac{1}{3}(4-r^2)^{3/2}\Big|_0^{2\cos\theta}\, d\theta \\
&= 2\int_0^{\pi/2} -\frac{8}{3}\sin^3\theta + \frac{8}{3}\, d\theta \\
&= 2\left(-\frac{8}{3}\frac{\cos^3\theta}{3} - \cos\theta + \frac{8}{3}\theta\right)\Big|_0^{\pi/2} \\
&= \frac{8}{3}\pi - \frac{32}{9}.
\end{aligned}$$

♣

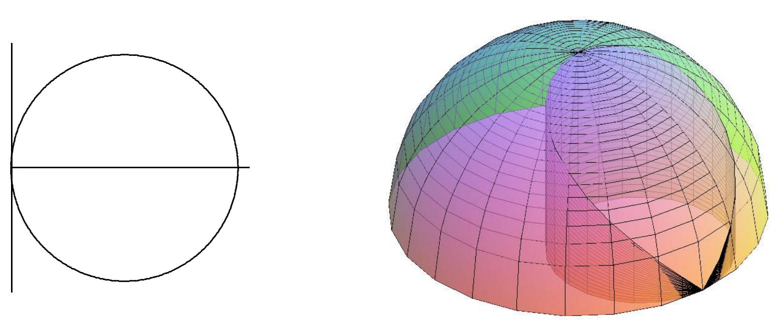


Figure 14.8: Volume over a region with non-constant limits.

You might have learned a formula for computing areas in polar coordinates. It is possible to compute areas as volumes, so that you need only remember one technique. Consider the surface $z = 1$, a horizontal plane. The volume under this surface and above a region in the x-y plane is simply $1 \cdot$ (area of the region), so computing the volume really just computes the area of the region.

Example 14.6:

Find the area outside the circle $r = 2$ and inside $r = 4\sin\theta$; see Figure 14.9.

Solution. The region is described by $\pi/6 \leq \theta \leq 5\pi/6$ and $2 \leq r \leq 4\sin\theta$, so the integral is

$$\begin{aligned}\int_{\pi/6}^{5\pi/6}\int_{2}^{4\sin\theta} 1\,r\,dr\,d\theta &= \int_{\pi/6}^{5\pi/6} \left.\frac{1}{2}r^2\right|_{2}^{4\sin\theta} d\theta \\ &= \int_{\pi/6}^{5\pi/6} 8\sin^2\theta - 2\,d\theta \\ &= \frac{4}{3}\pi + 2\sqrt{3}.\end{aligned}$$

♣

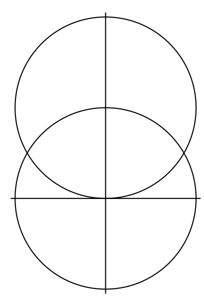


Figure 14.9: Finding area by computing volume.

Exercises for 14.2

Exercise 14.2.1 *Find the volume above the x-y plane, under the surface $r^2 = 2z$, and inside $r = 2$.*

Exercise 14.2.2 *Find the volume inside both $r = 1$ and $r^2 + z^2 = 4$.*

Exercise 14.2.3 *Find the volume below $z = \sqrt{1 - r^2}$ and above the top half of the cone $z = r$.*

Exercise 14.2.4 *Find the volume below $z = r$, above the x-y plane, and inside $r = \cos\theta$.*

Exercise 14.2.5 *Find the volume below $z = r$, above the x-y plane, and inside $r = 1 + \cos\theta$.*

Exercise 14.2.6 *Find the volume between* $x^2+y^2=z^2$ *and* $x^2+y^2=z$.

Exercise 14.2.7 *Find the area inside* $r=1+\sin\theta$ *and outside* $r=2\sin\theta$.

Exercise 14.2.8 *Find the area inside both* $r=2\sin\theta$ *and* $r=2\cos\theta$.

Exercise 14.2.9 *Find the area inside the four-leaf rose* $r=\cos(2\theta)$ *and outside* $r=1/2$.

Exercise 14.2.10 *Find the area inside the cardioid* $r=2(1+\cos\theta)$ *and outside* $r=2$.

Exercise 14.2.11 *Find the area of one loop of the three-leaf rose* $r=\cos(3\theta)$.

Exercise 14.2.12 *Compute* $\displaystyle\int_{-3}^{3}\int_{0}^{\sqrt{9-x^2}}\sin(x^2+y^2)\,dy\,dx$ *by converting to polar coordinates.*

Exercise 14.2.13 *Compute* $\displaystyle\int_{0}^{a}\int_{-\sqrt{a^2-x^2}}^{0}x^2y\,dy\,dx$ *by converting to polar coordinates.*

Exercise 14.2.14 *Find the volume under* $z=y^2+x+2$ *above the region* $x^2+y^2\le 4$

Exercise 14.2.15 *Find the volume between* $z=x^2y^3$ *and* $z=1$ *above the region* $x^2+y^2\le 1$

Exercise 14.2.16 *Find the volume inside* $x^2+y^2=1$ *and* $x^2+z^2=1$.

Exercise 14.2.17 *Find the volume under* $z=r$ *above* $r=3+\cos\theta$.

Exercise 14.2.18 *Figure 14.10 shows the plot of* $r=1+4\sin(5\theta)$.

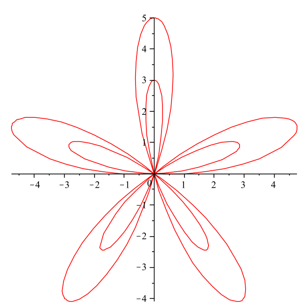


Figure 14.10: $r=1+4\sin(5\theta)$

(a) *Describe the behavior of the graph in terms of the given equation. Specifically, explain maximum and minimum values, number of leaves, and the 'leaves within leaves'.*

(b) *Give an integral or integrals to determine the area outside a smaller leaf but inside a larger leaf.*

(c) *How would changing the value of* a *in the equation* $r=1+a\cos(5\theta)$ *change the relative sizes of the inner and outer leaves? Focus on values* $a\ge 1$. *(Hint: How would we change the maximum and minimum values?)*

Exercise 14.2.19 *Consider the integral* $\iint_D \frac{1}{\sqrt{x^2+y^2}}\,dA$*, where* D *is the unit disk centered at the origin.*

(a) Why might this integral be considered improper?

(b) Calculate the value of the integral of the same function $1/\sqrt{x^2+y^2}$ *over the annulus with outer radius 1 and inner radius* δ*.*

(c) Obtain a value for the integral on the whole disk by letting δ *approach 0.*

(d) For which values λ *can we replace the denominator with* $(x^2+y^2)^\lambda$ *in the original integral?*

14.3 Moment and Center of Mass

Using a single integral we were able to compute the center of mass for a one-dimensional object with variable density, and a two dimensional object with constant density. With a double integral we can handle two dimensions and variable density.

Just as before, the coordinates of the center of mass are

$$\bar{x} = \frac{M_y}{M} \qquad \bar{y} = \frac{M_x}{M},$$

where M is the total mass, M_y is the moment around the y-axis, and M_x is the moment around the x-axis. (You may want to review the concepts in Section 8.6.)

The key to the computation, just as before, is the approximation of mass. In the two-dimensional case, we treat density σ as mass per square area, so when density is constant, mass is (density)(area). If we have a two-dimensional region with varying density given by $\sigma(x,y)$, and we divide the region into small subregions with area ΔA, then the mass of one subregion is approximately $\sigma(x_i,y_j)\Delta A$, the total mass is approximately the sum of many of these, and as usual the sum turns into an integral in the limit:

$$M = \int_{x_0}^{x_1} \int_{y_0}^{y_1} \sigma(x,y)\,dy\,dx,$$

and similarly for computations in cylindrical coordinates. Then as before

$$M_x = \int_{x_0}^{x_1} \int_{y_0}^{y_1} y\sigma(x,y)\,dy\,dx$$
$$M_y = \int_{x_0}^{x_1} \int_{y_0}^{y_1} x\sigma(x,y)\,dy\,dx.$$

Example 14.7: Center of Mass of Uniform Plate

Find the center of mass of a thin, uniform plate whose shape is the region between $y = \cos x$ *and the* x*-axis between* $x = -\pi/2$ *and* $x = \pi/2$*.*

Solution. Since the density is constant, we may take $\sigma(x,y) = 1$.

It is clear that $\bar{x} = 0$, but for practice let's compute it anyway. First we compute the mass:

$$M = \int_{-\pi/2}^{\pi/2} \int_0^{\cos x} 1\, dy\, dx = \int_{-\pi/2}^{\pi/2} \cos x\, dx = \sin x\Big|_{-\pi/2}^{\pi/2} = 2.$$

Next,

$$M_x = \int_{-\pi/2}^{\pi/2} \int_0^{\cos x} y\, dy\, dx = \int_{-\pi/2}^{\pi/2} \frac{1}{2}\cos^2 x\, dx = \frac{\pi}{4}.$$

Finally,

$$M_y = \int_{-\pi/2}^{\pi/2} \int_0^{\cos x} x\, dy\, dx = \int_{-\pi/2}^{\pi/2} x \cos x\, dx = 0.$$

So $\bar{x} = 0$ as expected, and $\bar{y} = \pi/4/2 = \pi/8$. This is the same problem as in Example 8.21; it may be helpful to compare the two solutions. ♣

Example 14.8: Center of Mass of 2-D Plate

Find the center of mass of a two-dimensional plate that occupies the quarter circle $x^2 + y^2 \leq 1$ in the first quadrant and has density $k(x^2 + y^2)$.

Solution. It seems clear that because of the symmetry of both the region and the density function (both are important!), $\bar{x} = \bar{y}$. We'll do both to check our work.

Jumping right in:

$$M = \int_0^1 \int_0^{\sqrt{1-x^2}} k(x^2 + y^2)\, dy\, dx = k \int_0^1 x^2\sqrt{1-x^2} + \frac{(1-x^2)^{3/2}}{3}\, dx.$$

This integral is something we can do, but it's a bit unpleasant. Since everything in sight is related to a circle, let's back up and try polar coordinates. Then $x^2 + y^2 = r^2$ and

$$M = \int_0^{\pi/2} \int_0^1 k(r^2)\, r\, dr\, d\theta = k \int_0^{\pi/2} \frac{r^4}{4}\Big|_0^1 d\theta = k \int_0^{\pi/2} \frac{1}{4}\, d\theta = k\frac{\pi}{8}.$$

Much better. Next, since $y = r\sin\theta$,

$$M_x = k \int_0^{\pi/2} \int_0^1 r^4 \sin\theta\, dr\, d\theta = k \int_0^{\pi/2} \frac{1}{5}\sin\theta\, d\theta = k -\frac{1}{5}\cos\theta\Big|_0^{\pi/2} = \frac{k}{5}.$$

Similarly,

$$M_y = k \int_0^{\pi/2} \int_0^1 r^4 \cos\theta\, dr\, d\theta = k \int_0^{\pi/2} \frac{1}{5}\cos\theta\, d\theta = k\,\frac{1}{5}\sin\theta\Big|_0^{\pi/2} = \frac{k}{5}.$$

Finally, $\bar{x} = \bar{y} = \dfrac{8}{5\pi}$. ♣

Exercises for 14.3

Exercise 14.3.1 *Find the center of mass of a two-dimensional plate that occupies the square* $[0,1]\times[0,1]$ *and has density function* xy.

Exercise 14.3.2 *Find the center of mass of a two-dimensional plate that occupies the triangle* $0\le x\le 1$, $0\le y\le x$, *and has density function* xy.

Exercise 14.3.3 *Find the center of mass of a two-dimensional plate that occupies the upper unit semicircle centered at* $(0,0)$ *and has density function* y.

Exercise 14.3.4 *Find the center of mass of a two-dimensional plate that occupies the upper unit semicircle centered at* $(0,0)$ *and has density function* x^2.

Exercise 14.3.5 *Find the center of mass of a two-dimensional plate that occupies the triangle formed by* $x=2$, $y=x$, *and* $y=2x$ *and has density function* $2x$.

Exercise 14.3.6 *Find the center of mass of a two-dimensional plate that occupies the triangle formed by* $x=0$, $y=x$, *and* $2x+y=6$ *and has density function* x^2.

Exercise 14.3.7 *Find the center of mass of a two-dimensional plate that occupies the region enclosed by the parabolas* $x=y^2$, $y=x^2$ *and has density function* $\sqrt{x}$.

Exercise 14.3.8 *Find the centroid of the area in the first quadrant bounded by* $x^2-8y+4=0$, $x^2=4y$, *and* $x=0$. *(Recall that the centroid is the center of mass when the density is 1 everywhere.)*

Exercise 14.3.9 *Find the centroid of one loop of the three-leaf rose* $r=\cos(3\theta)$. *(Recall that the centroid is the center of mass when the density is 1 everywhere, and that the mass in this case is the same as the area, which was the subject of Exercise 14.2.11 in Section 14.2.) The computations of the integrals for the moments* M_x *and* M_y *are elementary but quite long; Sage can help.*

Exercise 14.3.10 *Find the center of mass of a two dimensional object that occupies the region* $0\le x\le \pi$, $0\le y\le \sin x$, *with density* $\sigma=1$.

Exercise 14.3.11 *A two-dimensional object has shape given by* $r=1+\cos\theta$ *and density* $\sigma(r,\theta)=2+\cos\theta$. *Set up the three integrals required to compute the center of mass.*

Exercise 14.3.12 *A two-dimensional object has shape given by* $r=\cos\theta$ *and density* $\sigma(r,\theta)=r+1$. *Set up the three integrals required to compute the center of mass.*

Exercise 14.3.13 *A two-dimensional object sits inside* $r=1+\cos\theta$ *and outside* $r=\cos\theta$, *and has density* 1 *everywhere. Set up the integrals required to compute the center of mass.*

14.4 Surface Area

We next seek to compute the area of a surface above (or below) a region in the x-y plane. How might we approximate this? We start, as usual, by dividing the region into a grid of small rectangles. We want to approximate the area of the surface above one of these small rectangles. The area is very close to the area of the tangent plane above the small rectangle. If the tangent plane just happened to be horizontal, of course the area would simply be the area of the rectangle. For a typical plane, however, the area is the area of a parallelogram, as indicated in Figure 14.11. Note that the area of the parallelogram is obviously larger the more "tilted" the tangent plane.

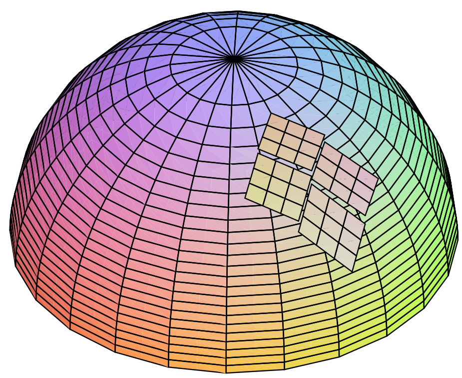


Figure 14.11: Small parallelograms at points of tangency.

Now recall a curious fact: the area of a parallelogram can be computed as the cross product of two vectors. We simply need to acquire two vectors, parallel to the sides of the parallelogram and with lengths to match. But this is easy: in the x direction we use the tangent vector we already know, namely $\langle 1, 0, f_x\rangle$ and multiply by Δx to shrink it to the right size: $\langle \Delta x, 0, f_x\Delta x\rangle$. In the y direction we do the same thing and get $\langle 0, \Delta y, f_y\Delta y\rangle$. The cross product of these vectors is $\langle f_x, f_y, -1\rangle\,\Delta x\Delta y$ with length $\sqrt{f_x^2+f_y^2+1}\,\Delta x\Delta y$, the area of the parallelogram. Now we add these up and take the limit, to produce the integral

$$\int_{x_0}^{x_1}\int_{y_0}^{y_1}\sqrt{f_x^2+f_y^2+1}\,dy\,dx.$$

As before, the limits need not be constant.

Example 14.9: Surface Area of a Hemisphere

Find the area of the hemisphere $z=\sqrt{1-x^2-y^2}$.

Solution. We compute the derivatives

$$f_x=\frac{-x}{\sqrt{1-x^2-y^2}}\qquad f_x=\frac{-y}{\sqrt{1-x^2-y^2}},$$

and then the area is

$$\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\sqrt{\frac{x^2}{1-x^2-y^2}+\frac{y^2}{1-x^2-y^2}+1}\,dy\,dx.$$

This is a bit on the messy side, but we can use polar coordinates:

$$\int_0^{2\pi}\int_0^1\sqrt{\frac{1}{1-r^2}}\,r\,dr\,d\theta.$$

This integral is improper, since the function is undefined at the limit 1. We therefore compute

$$\lim_{a\to 1^-}\int_0^a\sqrt{\frac{1}{1-r^2}}\,r\,dr=\lim_{a\to 1^-}-\sqrt{1-a^2}+1=1,$$

using the substitution $u=1-r^2$. Then the area is

$$\int_0^{2\pi}1\,d\theta=2\pi.$$

You may recall that the area of a sphere of radius r is $4\pi r^2$, so half the area of a unit sphere is $(1/2)4\pi=2\pi$, in agreement with our answer. ♣

Exercises for 14.4

Exercise 14.4.1 *Find the area of the surface of a right circular cone of height h and base radius a.*

Exercise 14.4.2 *Find the area of the portion of the plane $z=mx$ inside the cylinder $x^2+y^2=a^2$.*

Exercise 14.4.3 *Find the area of the portion of the plane $x+y+z=1$ in the first octant.*

Exercise 14.4.4 *Find the area of the upper half of the cone $x^2+y^2=z^2$ inside the cylinder $x^2+y^2-2x=0$.*

Exercise 14.4.5 *Find the area of the upper half of the cone $x^2+y^2=z^2$ above the interior of one loop of $r=\cos(2\theta)$.*

Exercise 14.4.6 *Find the area of the upper hemisphere of $x^2+y^2+z^2=1$ above the interior of one loop of $r=\cos(2\theta)$.*

Exercise 14.4.7 *The plane $ax+by+cz=d$ cuts a triangle in the first octant provided that a,b,c and d are all positive. Find the area of this triangle.*

Exercise 14.4.8 *Find the area of the portion of the cone $x^2+y^2=3z^2$ lying above the xy plane and inside the cylinder $x^2+y^2=4y$.*

14.5 Triple Integrals

It will come as no surprise that we can also do triple integrals—integrals over a three-dimensional region. The simplest application allows us to compute volumes in an alternate way.

To approximate a volume in three dimensions, we can divide the three-dimensional region into small rectangular boxes, each $\Delta x \times \Delta y \times \Delta z$ with volume $\Delta x \Delta y \Delta z$. Then we add them all up and take the limit, to get an integral:

$$\int_{x_0}^{x_1} \int_{y_0}^{y_1} \int_{z_0}^{z_1} dz\, dy\, dx.$$

Of course, if the limits are constant, we are simply computing the volume of a rectangular box.

Example 14.10: Volume of a Box

Compute the volume of the box with opposite corners at $(0,0,0)$ *and* $(1,2,3)$.

Solution. We use an integral to compute the volume of the box:

$$\int_0^1 \int_0^2 \int_0^3 dz\, dy\, dx = \int_0^1 \int_0^2 z\big|_0^3 \, dy\, dx = \int_0^1 \int_0^2 3\, dy\, dx = \int_0^1 3y\big|_0^2 \, dx = \int_0^1 6\, dx = 6.$$

♣

Of course, this is more interesting and useful when the limits are not constant.

Example 14.11: Volume of a Tetrahedron

Find the volume of the tetrahedron with corners at $(0,0,0)$, $(0,3,0)$, $(2,3,0)$, *and* $(2,3,5)$.

Solution. The whole problem comes down to correctly describing the region by inequalities: $0 \le x \le 2$, $3x/2 \le y \le 3$, $0 \le z \le 5x/2$. The lower y limit comes from the equation of the line $y = 3x/2$ that forms one edge of the tetrahedron in the x-y plane; the upper z limit comes from the equation of the plane $z = 5x/2$ that forms the "upper" side of the tetrahedron; see Figure 14.12. Now the volume is

$$\begin{aligned}
\int_0^2 \int_{3x/2}^3 \int_0^{5x/2} dz\, dy\, dx &= \int_0^2 \int_{3x/2}^3 z\big|_0^{5x/2} \, dy\, dx \\
&= \int_0^2 \int_{3x/2}^3 \frac{5x}{2}\, dy\, dx \\
&= \int_0^2 \left.\frac{5x}{2} y\right|_{3x/2}^3 dx \\
&= \int_0^2 \frac{15x}{2} - \frac{15x^2}{4}\, dx \\
&= \left.\frac{15x^2}{4} - \frac{15x^3}{12}\right|_0^2
\end{aligned}$$

$$= 15 - 10 = 5.$$

♣

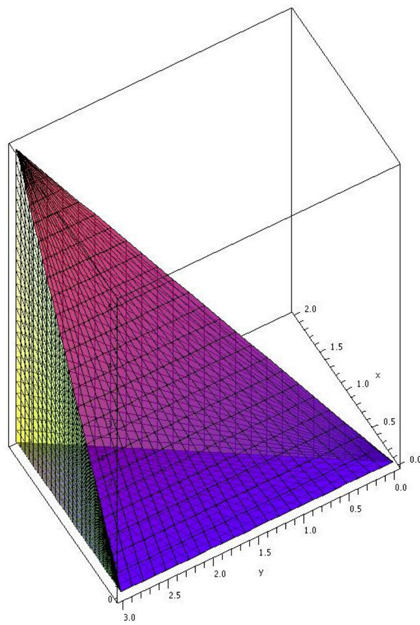


Figure 14.12: A tetrahedron.

Pretty much just the way we did for two dimensions we can use triple integration to compute mass, center of mass, and various average quantities.

Example 14.12: Average Temperature in a Cube

Suppose the temperature at a point is given by $T = xyz$. Find the average temperature in the cube with opposite corners at $(0,0,0)$ and $(2,2,2)$.

Solution. In two dimensions we add up the temperature at "each" point and divide by the area; here we add up the temperatures and divide by the volume, 8:

$$\begin{aligned}\frac{1}{8}\int_0^2\int_0^2\int_0^2 xyz\,dz\,dy\,dx &= \frac{1}{8}\int_0^2\int_0^2 \left.\frac{xyz^2}{2}\right|_0^2 dy\,dx = \frac{1}{16}\int_0^2\int_0^2 xy\,dy\,dx \\ &= \frac{1}{4}\int_0^2 \left.\frac{xy^2}{2}\right|_0^2 dx = \frac{1}{8}\int_0^2 4x\,dx = \left.\frac{1}{2}\,\frac{x^2}{2}\right|_0^2 = 1.\end{aligned}$$

♣

Example 14.13: Mass & Center of Mass of Tetrahedron

Suppose the density of an object is given by xz, and the object occupies the tetrahedron with corners $(0,0,0)$, $(0,1,0)$, $(1,1,0)$, and $(0,1,1)$. Find the mass and center of mass of the object.

Solution. As usual, the mass is the integral of density over the region:

$$M = \int_0^1 \int_x^1 \int_0^{y-x} xz\,dz\,dy\,dx = \int_0^1 \int_x^1 \frac{x(y-x)^2}{2}\,dy\,dx = \frac{1}{2}\int_0^1 \frac{x(1-x)^3}{3}\,dx$$
$$= \frac{1}{6}\int_0^1 x - 3x^2 + 3x^3 - x^4\,dx = \frac{1}{120}.$$

We compute moments as before, except now there is a third moment:

$$M_{xy} = \int_0^1 \int_x^1 \int_0^{y-x} xz^2\,dz\,dy\,dx = \frac{1}{360},$$
$$M_{xz} = \int_0^1 \int_x^1 \int_0^{y-x} xyz\,dz\,dy\,dx = \frac{1}{144},$$
$$M_{yz} = \int_0^1 \int_x^1 \int_0^{y-x} x^2z\,dz\,dy\,dx = \frac{1}{360}.$$

Finally, the coordinates of the center of mass are $\bar{x} = M_{yz}/M = 1/3$, $\bar{y} = M_{xz}/M = 5/6$, and $\bar{z} = M_{xy}/M = 1/3$. ♣

Exercises for 14.5

Exercise 14.5.1 *Evaluate* $\displaystyle\int_0^1 \int_0^x \int_0^{x+y} 2x + y - 1\,dz\,dy\,dx.$

Exercise 14.5.2 *Evaluate* $\displaystyle\int_0^2 \int_{-1}^{x^2} \int_1^y xyz\,dz\,dy\,dx.$

Exercise 14.5.3 *Evaluate* $\displaystyle\int_0^1 \int_0^x \int_0^{\ln y} e^{x+y+z}\,dz\,dy\,dx.$

Exercise 14.5.4 *Evaluate* $\displaystyle\int_0^{\pi/2} \int_0^{\sin\theta} \int_0^{r\cos\theta} r^2\,dz\,dr\,d\theta.$

Exercise 14.5.5 *Evaluate* $\displaystyle\int_0^{\pi} \int_0^{\sin\theta} \int_0^{r\sin\theta} r\cos^2\theta\,dz\,dr\,d\theta.$

Exercise 14.5.6 *Evaluate* $\displaystyle\int_0^1 \int_0^{y^2} \int_0^{x+y} x\,dz\,dx\,dy.$

Exercise 14.5.7 *Evaluate* $\displaystyle\int_1^2 \int_y^{y^2} \int_0^{\ln(y+z)} e^x\,dx\,dz\,dy.$

Exercise 14.5.8 *Compute* $\displaystyle\int_0^{\pi} \int_0^{\pi/2} \int_0^1 z\sin x + z\cos y\,dz\,dy\,dx.$

Exercise 14.5.9 *For each of the integrals in the previous exercises, give a description of the volume (both algebraic and geometric) that is the domain of integration.*

Exercise 14.5.10 *Compute* $\int\int\int x+y+z\,dV$ *over the region inside* $x^2+y^2+z^2 \leq 1$ *in the first octant.*

Exercise 14.5.11 *Find the mass of a cube with edge length 2 and density equal to the square of the distance from one corner.*

Exercise 14.5.12 *Find the mass of a cube with edge length 2 and density equal to the square of the distance from one edge.*

Exercise 14.5.13 *An object occupies the volume of the upper hemisphere of* $x^2+y^2+z^2=4$ *and has density* z *at* (x,y,z)*. Find the center of mass.*

Exercise 14.5.14 *An object occupies the volume of the pyramid with corners at* $(1,1,0)$, $(1,-1,0)$, $(-1,-1,0)$, $(-1,1,0)$, *and* $(0,0,2)$ *and has density* x^2+y^2 *at* (x,y,z)*. Find the center of mass.*

Exercise 14.5.15 *Verify the moments* M_{xy}, M_{xz}, *and* M_{yz} *of Example 14.13 by evaluating the integrals.*

Exercise 14.5.16 *Find the region* E *for which* $\iiint_E (1-x^2-y^2-z^2)\,dV$ *is a maximum.*

14.6 Cylindrical and Spherical Coordinates

We have seen that sometimes double integrals are simplified by doing them in polar coordinates; not surprisingly, triple integrals are sometimes simpler in cylindrical coordinates or spherical coordinates. To set up integrals in polar coordinates, we had to understand the shape and area of a typical small region into which the region of integration was divided. We need to do the same thing here, for three dimensional regions.

The cylindrical coordinate system is the simplest, since it is just the polar coordinate system plus a z coordinate. A typical small unit of volume is the shape shown in Figure 14.7 “fattened up” in the z direction, so its volume is $r\Delta r\Delta\theta\Delta z$, or in the limit, $r\,dr\,d\theta\,dz$.

Example 14.14: Finding Volume

Find the volume under $z=\sqrt{4-r^2}$ *above the quarter circle inside* $x^2+y^2=4$ *in the first quadrant.*

Solution. We could of course do this with a double integral, but we’ll use a triple integral:

$$\int_0^{\pi/2}\int_0^2\int_0^{\sqrt{4-r^2}} r\,dz\,dr\,d\theta = \int_0^{\pi/2}\int_0^2 \sqrt{4-r^2}\,r\,dr\,d\theta = \frac{4\pi}{3}.$$

Compare this to Example 14.5. ♣

Example 14.15: Mass using Cylindrical Coordinates

An object occupies the space inside both the cylinder $x^2+y^2=1$ *and the sphere* $x^2+y^2+z^2=4$*, and has density* x^2 *at* (x,y,z)*. Find the total mass.*

Solution. We set this up in cylindrical coordinates, recalling that $x = r\cos\theta$:

$$\begin{aligned}\int_0^{2\pi}\int_0^1\int_{-\sqrt{4-r^2}}^{\sqrt{4-r^2}} r^3\cos^2(\theta)\,dz\,dr\,d\theta &= \int_0^{2\pi}\int_0^1 2\sqrt{4-r^2}\,r^3\cos^2(\theta)\,dr\,d\theta \\ &= \int_0^{2\pi}\left(\frac{128}{15}-\frac{22}{5}\sqrt{3}\right)\cos^2(\theta)\,d\theta \\ &= \left(\frac{128}{15}-\frac{22}{5}\sqrt{3}\right)\pi\end{aligned}$$

♣

Spherical coordinates are somewhat more difficult to understand. The small volume we want will be defined by $\Delta\rho$, $\Delta\phi$, and $\Delta\theta$, as pictured in Figure 14.13. The small volume is nearly box shaped, with 4 flat sides and two sides formed from bits of concentric spheres. When $\Delta\rho$, $\Delta\phi$, and $\Delta\theta$ are all very small, the volume of this little region will be nearly the volume we get by treating it as a box. One dimension of the box is simply $\Delta\rho$, the change in distance from the origin. The other two dimensions are the lengths of small circular arcs, so they are $r\Delta\alpha$ for some suitable r and α, just as in the polar coordinates case.

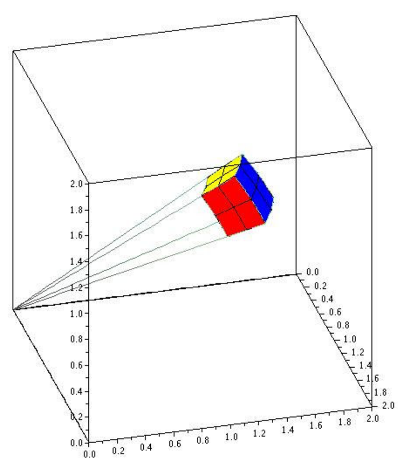


Figure 14.13: A small unit of volume for spherical coordinates.

The easiest of these to understand is the arc corresponding to a change in ϕ, which is nearly identical to the derivation for polar coordinates, as shown in the left graph in Figure 14.14. In that graph we are looking "face on" at the side of the box we are interested in, so the small angle pictured is precisely $\Delta\phi$, the vertical axis really is the z axis, but the horizontal axis is *not* a real axis—it is just some line in the x-y plane. Because the other arc is governed by θ, we need to imagine looking straight down the z axis, so that the apparent angle we see is $\Delta\theta$. In this view, the axes really are the x and y axes. In this graph, the apparent distance from the origin is not ρ but $\rho\sin\phi$, as indicated in the left graph.

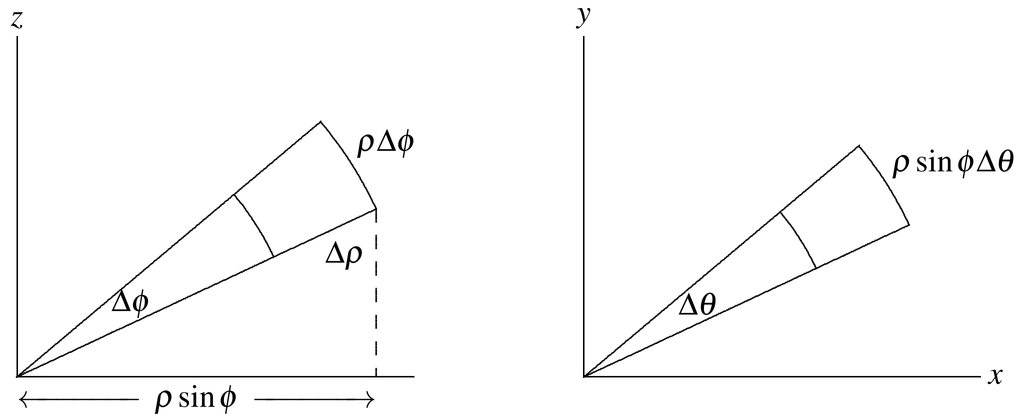


Figure 14.14: Setting up integration in spherical coordinates.

The upshot is that the volume of the little box is approximately $\Delta\rho(\rho\Delta\phi)(\rho\sin\phi\Delta\theta) = \rho^2\sin\phi\Delta\rho\Delta\phi\Delta\theta$, or in the limit $\rho^2\sin\phi\,d\rho\,d\phi\,d\theta$.

Example 14.16: Average Temperature in a Unit Sphere

Suppose the temperature at (x,y,z) is $T = 1/(1+x^2+y^2+z^2)$. Find the average temperature in the unit sphere centered at the origin.

Solution. In two dimensions we add up the temperature at "each" point and divide by the area; here we add up the temperatures and divide by the volume, $(4/3)\pi$:

$$\frac{3}{4\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}}\frac{1}{1+x^2+y^2+z^2}\,dz\,dy\,dx$$

This looks quite messy; since everything in the problem is closely related to a sphere, we'll convert to spherical coordinates.

$$\frac{3}{4\pi}\int_{0}^{2\pi}\int_{0}^{\pi}\int_{0}^{1}\frac{1}{1+\rho^2}\rho^2\sin\phi\,d\rho\,d\phi\,d\theta = \frac{3}{4\pi}(4\pi-\pi^2) = 3-\frac{3\pi}{4}.$$

♣

Exercises for 14.6

Exercise 14.6.1 *Evaluate* $\displaystyle\int_0^1\int_0^x\int_0^{\sqrt{x^2+y^2}}\frac{(x^2+y^2)^{3/2}}{x^2+y^2+z^2}\,dz\,dy\,dx.$

Exercise 14.6.2 *Evaluate* $\displaystyle\int_{-1}^{1}\int_{0}^{\sqrt{1-x^2}}\int_{\sqrt{x^2+y^2}}^{\sqrt{2-x^2-y^2}} \sqrt{x^2+y^2+z^2}\,dz\,dy\,dx.$

Exercise 14.6.3 *Evaluate* $\displaystyle\int\int\int x^2\,dV$ *over the interior of the cylinder* $x^2+y^2=1$ *between* $z=0$ *and* $z=5$.

Exercise 14.6.4 *Evaluate* $\displaystyle\int\int\int xy\,dV$ *over the interior of the cylinder* $x^2+y^2=1$ *between* $z=0$ *and* $z=5$.

Exercise 14.6.5 *Evaluate* $\displaystyle\int\int\int z\,dV$ *over the region above the x-y plane, inside* $x^2+y^2-2x=0$ *and under* $x^2+y^2+z^2=4$.

Exercise 14.6.6 *Evaluate* $\displaystyle\int\int\int yz\,dV$ *over the region in the first octant, inside* $x^2+y^2-2x=0$ *and under* $x^2+y^2+z^2=4$.

Exercise 14.6.7 *Evaluate* $\displaystyle\int\int\int x^2+y^2\,dV$ *over the interior of* $x^2+y^2+z^2=4$.

Exercise 14.6.8 *Evaluate* $\displaystyle\int\int\int \sqrt{x^2+y^2}\,dV$ *over the interior of* $x^2+y^2+z^2=4$.

Exercise 14.6.9 *Compute* $\displaystyle\int\int\int x+y+z\,dV$ *over the region inside* $x^2+y^2+z^2=1$ *in the first octant.*

Exercise 14.6.10 *Find the mass of a right circular cone of height h and base radius a if the density is proportional to the distance from the base.*

Exercise 14.6.11 *Find the mass of a right circular cone of height h and base radius a if the density is proportional to the distance from its axis of symmetry.*

Exercise 14.6.12 *An object occupies the region inside the unit sphere at the origin, and has density equal to the distance from the x-axis. Find the mass.*

Exercise 14.6.13 *An object occupies the region inside the unit sphere at the origin, and has density equal to the square of the distance from the origin. Find the mass.*

Exercise 14.6.14 *An object occupies the region between the unit sphere at the origin and a sphere of radius 2 with center at the origin, and has density equal to the distance from the origin. Find the mass.*

Exercise 14.6.15 *An object occupies the region in the first octant bounded by the cones* $\phi=\pi/4$ *and* $\phi=\arctan 2$, *and the sphere* $\rho=\sqrt{6}$, *and has density proportional to the distance from the origin. Find the mass.*

14.7 Change of Variables

One of the most useful techniques for evaluating integrals is substitution, both "u-substitution" and trigonometric substitution, in which we change the variable to something more convenient. As we have seen, sometimes changing from rectangular coordinates to another coordinate system is helpful, and this too changes the variables. This is certainly a more complicated change, since instead of changing one variable for another we change an entire suite of variables, but as it turns out it is really very similar to the kinds of change of variables we already know as substitution.

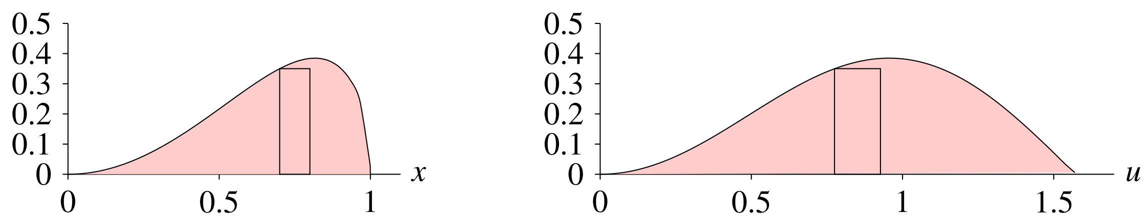


Figure 14.15: Single change of variable.

Let's examine the single variable case again, from a slightly different perspective than we have previously used. Suppose we start with the problem

$$\int_0^1 x^2\sqrt{1-x^2}\,dx;$$

this computes the area in the left graph of Figure 14.15. We use the substitution $x=\sin u$ to transform the function from $x^2\sqrt{1-x^2}$ to $\sin^2 u\sqrt{1-\sin^2 u}$, and we also convert dx to $\cos u\,du$. Finally, we convert the limits 0 and 1 to 0 and $\pi/2$. This transforms the integral:

$$\int_0^1 x^2\sqrt{1-x^2}\,dx = \int_0^{\pi/2} \sin^2 u\sqrt{1-\sin^2 u}\cos u\,du.$$

We want to notice that there are three different conversions: the main function, the differential dx, and the interval of integration. The function is converted to $\sin^2 u\sqrt{1-\sin^2 u}$, shown in the right-hand graph of Figure 14.15. It is evident that the two curves pictured there have the same y-values in the same order, but the horizontal scale has been changed. Even though the heights are the same, the two integrals

$$\int_0^1 x^2\sqrt{1-x^2}\,dx \qquad \text{and} \qquad \int_0^{\pi/2} \sin^2 u\sqrt{1-\sin^2 u}\,du$$

are not the same; clearly the right hand area is larger. One way to understand the problem is to note that if both areas are approximated using, say, ten subintervals, that the approximating rectangles on the right are wider than their counterparts on the left, as indicated. In the picture, the width of the rectangle on the left is $\Delta x = 0.1$, between 0.7 and 0.8. The rectangle on the right is situated between the corresponding values $\arcsin(0.7)$ and $\arcsin(0.8)$ so that $\Delta u = \arcsin(0.8)-\arcsin(0.7)$. To make the widths match, and the areas therefore the same, we can multiply Δu by a correction factor; in this case the correction factor is approximately $\cos u = \cos(\arcsin(0.7))$, which we compute when we convert dx to $\cos u\,du$.

Now let's move to functions of two variables. Suppose we want to convert an integral

$$\int_{x_0}^{x_1}\int_{y_0}^{y_1} f(x,y)\,dy\,dx$$

to use new variables u and v. In the single variable case, there's typically just one reason to want to change the variable: to make the function "nicer" so that we can find an antiderivative. In the two variable case, there is a second potential reason: the two-dimensional region over which we need to integrate is somehow unpleasant, and we want the region in terms of u and v to be nicer—to be a rectangle, for example. Ideally, of course, the new function and the new region will be no worse than the originals, and at least one of them will be better; this doesn't always pan out.

As before, there are three parts to the conversion: the function itself must be rewritten in terms of u and v, $dy\,dx$ must be converted to $du\,dv$, and the old region must be converted to the new region. We will develop the necessary techniques by considering a particular example, and we will use an example we already know how to do by other means.

Consider

$$\int_{-1}^{1}\int_{0}^{\sqrt{1-x^2}} \sqrt{x^2+y^2}\,dy\,dx.$$

The limits correspond to integrating over the top half of a circular disk, and we recognize that the function will simplify in polar coordinates, so we would normally convert to polar coordinates:

$$\int_0^{\pi}\int_0^1 \sqrt{r^2}\,r\,dr\,d\theta = \frac{\pi}{3}.$$

But let's instead approach this as a substitution problem, starting with $x = r\cos\theta$, $y = r\sin\theta$. This pair of equations describes a function from "r-θ space" to "x-y space", and because it involves familiar concepts, it is not too hard to understand what it does. In Figure 14.16 we have indicated geometrically a bit about how this function behaves. The four dots labeled a–d in the r-θ plane correspond to the three dots in the x-y plane; dots a and b both go to the origin because $r = 0$. The horizontal arrow in the r-θ plane has $r = 1$ everywhere and θ ranges from 0 to π, so the corresponding points $x = r\cos\theta$, $y = r\sin\theta$ start at $(1,0)$ and follow the unit circle counter-clockwise. Finally, the vertical arrow has $\theta = \pi/4$ and r ranges from 0 to 1, so it maps to the straight arrow in the x-y plane. Extrapolating from these few examples, it's not hard to see that every vertical line in the r-θ plane is transformed to a line through the origin in the x-y plane, and every horizontal line in the r-θ plane is transformed to a circle with center at the origin in the x-y plane. Since we are interested in integrating over the half-disk in the x-y plane, we will integrate over the rectangle $[0,\pi]\times[0,1]$ in the r-θ plane, because we now see that the points in this rectangle are sent precisely to the upper half disk by $x = r\cos\theta$ and $y = r\sin\theta$.

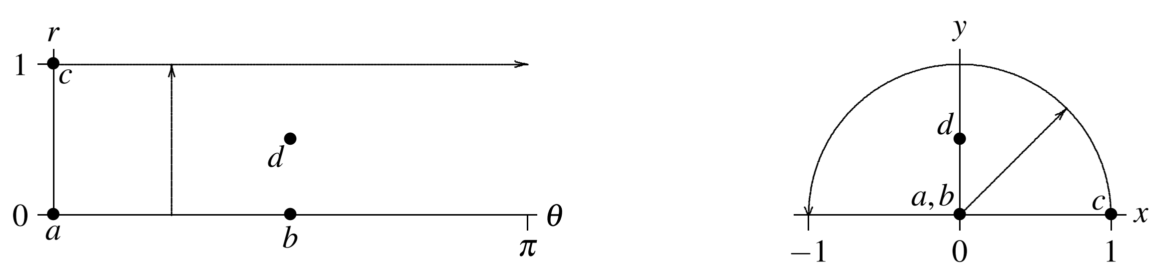


Figure 14.16: Double change of variable.

At this point we are two-thirds done with the task: we know the r-θ limits of integration, and we can easily convert the function to the new variables:

$$\sqrt{x^2+y^2} = \sqrt{r^2\cos^2\theta + r^2\sin^2\theta} = r\sqrt{\cos^2\theta+\sin^2\theta} = r. \tag{14.1}$$

The final, and most difficult, task is to figure out what replaces $dx\,dy$. (Of course, we actually know the answer, because we are in effect converting to polar coordinates. What we really want is a series of steps that gets to that right answer but that will also work for other substitutions that are not so familiar.)

Let's take a step back and remember how integration arises from approximation. When we approximate the integral in the x-y plane, we are computing the volumes of tall thin boxes, in this case boxes that are $\Delta x \times \Delta y \times \sqrt{x^2+y^2}$. We are aiming to come up with an integral in the r-θ plane that looks like this:

$$\int_0^\pi \int_0^1 r(?)\,dr\,d\theta. \tag{14.2}$$

What we're missing is exactly the right quantity to replace the "?" so that we get the correct answer. Of course, this integral is also the result of an approximation, in which we add up volumes of boxes that are $\Delta r \times \Delta\theta \times$ height; the problem is that the height that will give us the correct answer is not simply r. Or put another way, we can think of the correct height as r, but the area of the base $\Delta r\Delta\theta$ as being wrong. The height r comes from Equation 14.1, which is to say, it is precisely the same as the corresponding height in the x-y version of the integral. The problem is that the area of the base $\Delta x \times \Delta y$ is not the same as the area of the base $\Delta r \times \Delta\theta$. We can think of the "?" in the integral as a correction factor that is needed so that $?\,dr\,d\theta = dx\,dy$.

So let's think about what that little base $\Delta r \times \Delta\theta$ corresponds to. We know that each bit of horizontal line in the r-θ plane corresponds to a bit of circular arc in the x-y plane, and each bit of vertical line in the r-θ plane corresponds to a bit of "radial line" in the x-y plane. In Figure 14.17 we show a typical rectangle in the r-θ plane and its corresponding area in the x-y plane.

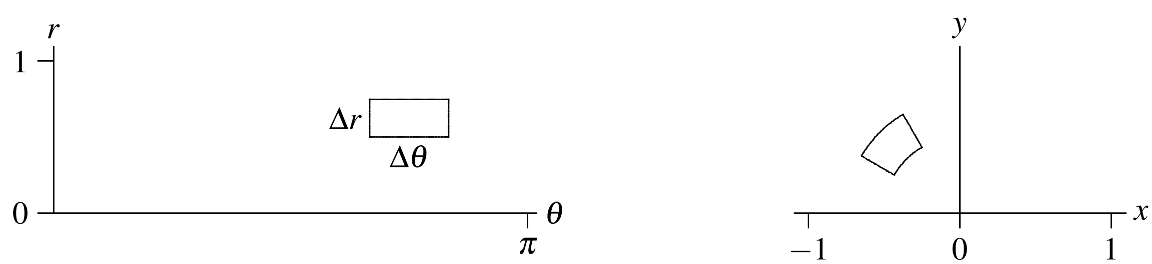


Figure 14.17: Corresponding areas.

In this case, the region in the x-y plane is approximately a rectangle with dimensions $\Delta r \times r\Delta\theta$, but in general the corner angles will not be right angles, so the region will typically be (almost) a parallelogram. We need to compute the area of this parallelogram. We know a neat way to do this: compute the length of a certain cross product. If we can determine an appropriate two vectors we'll be nearly done.

Fortunately, we've really done this before. The sides of the region in the x-y plane are formed by temporarily fixing either r or θ and letting the other variable range over a small interval. In Figure 14.17, for example, the upper right edge of the region is formed by fixing $\theta = 2\pi/3$ and letting r run from 0.5 to 0.75. In other words, we have a vector function $\mathbf{v}(r) = \langle r\cos\theta_0, r\sin\theta_0, 0\rangle$, and we are interested in a restricted set of values for r. A vector tangent to this path is given by the derivative $\mathbf{v}'(r) = \langle \cos\theta_0, \sin\theta_0, 0\rangle$, and a small tangent vector, with length approximately equal to the side of the region, is $\langle \cos\theta_0, \sin\theta_0, 0\rangle\,dr$. Likewise, if we fix $r = r_0 = 0.5$, we get the vector function $\mathbf{w}(\theta) = \langle r_0\cos\theta, r_0\sin\theta, 0\rangle$ with derivative $\mathbf{w}(\theta) = \langle -r_0\sin\theta, r_0\cos\theta, 0\rangle$ and a small tangent vector $\langle -r_0\sin\theta_0, r_0\cos\theta_0, 0\rangle\,d\theta$ when $\theta = \theta_0$ (at the corner we're focusing on). These vectors are shown in Figure 14.18, with the actual region outlined by a dotted boundary. Of course, since both Δr and $\Delta\theta$ are quite large, the parallelogram is not a particularly good approximation to the true area.

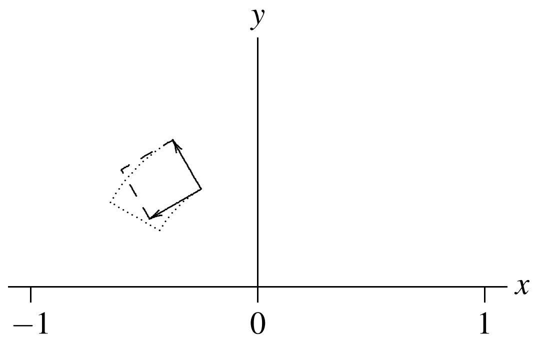


Figure 14.18: The approximating parallelogram.

The area of this parallelogram is the length of the cross product:

$$\begin{aligned}\langle -r_0\sin\theta_0, r_0\cos\theta_0, 0\rangle\,d\theta \times \langle \cos\theta_0, \sin\theta_0, 0\rangle\,dr &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -r_0\sin\theta_0 & r_0\cos\theta_0 & 0 \\ \cos\theta_0 & \sin\theta_0 & 0 \end{vmatrix} d\theta\,dr \\ &= \langle 0, 0, -r_0\sin^2\theta_0 - r_0\cos^2\theta_0\rangle\,d\theta\,dr \\ &= \langle 0, 0, -r_0\rangle\,d\theta\,dr.\end{aligned}$$

The length of this vector is $r_0\,dr\,d\theta$. So in general, for any values of r and θ, the area in the x-y plane corresponding to a small rectangle anchored at (θ, r) in the r-θ plane is approximately $r\,dr\,d\theta$. In other words, "r" replaces the "?" in Equation 14.2.

In general, a substitution will start with equations $x = f(u,v)$ and $y = g(u,v)$. Again, it will be straightforward to convert the function being integrated. Converting the limits will require, as above, an understanding of just how the functions f and g transform the u-v plane into the x-y plane. Finally, the small vectors we need to approximate an area will be $\langle f_u, g_u, 0\rangle\,du$ and $\langle f_v, g_v, 0\rangle\,dv$. The cross product of these is $\langle 0, 0, f_u g_v - g_u f_v\rangle\,du\,dv$ with length $|f_u g_v - g_u f_v|\,du\,dv$. The quantity $|f_u g_v - g_u f_v|$ is usually denoted

$$\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = |f_u g_v - g_u f_v|$$

and called the **Jacobian**. Note that this is the absolute value of the two by two determinant

$$\begin{vmatrix} f_u & g_u \\ f_v & g_v \end{vmatrix},$$

which may be easier to remember. (Confusingly, the matrix, the determinant of the matrix, and the absolute value of the determinant are all called the Jacobian by various authors.)

Because there are two things to worry about, namely, the form of the function and the region of integration, transformations in two (or more) variables are quite tricky to discover.

Example 14.17: Integral of an Ellipse

Integrate $x^2 - xy + y^2$ over the region $x^2 - xy + y^2 \le 2$.

Solution. The equation $x^2 - xy + y^2 = 2$ describes an ellipse as in Figure 14.19; the region of integration is the interior of the ellipse. We will use the transformation $x = \sqrt{2}u - \sqrt{2/3}v$, $y = \sqrt{2}u + \sqrt{2/3}v$. Substituting into the function itself we get

$$x^2 - xy + y^2 = 2u^2 + 2v^2.$$

The boundary of the ellipse is $x^2 - xy + y^2 = 2$, so the boundary of the corresponding region in the u-v plane is $2u^2 + 2v^2 = 2$ or $u^2 + v^2 = 1$, the unit circle, so this substitution makes the region of integration simpler.

Next, we compute the Jacobian, using $f = \sqrt{2}u - \sqrt{2/3}v$ and $g = \sqrt{2}u + \sqrt{2/3}v$:

$$f_u g_v - g_u f_v = \sqrt{2}\sqrt{2/3} + \sqrt{2}\sqrt{2/3} = \frac{4}{\sqrt{3}}.$$

Hence the new integral is

$$\iint_R (2u^2 + 2v^2)\frac{4}{\sqrt{3}}\,du\,dv,$$

where R is the interior of the unit circle. This is still not an easy integral, but it is easily transformed to polar coordinates, and then easily integrated. ♣

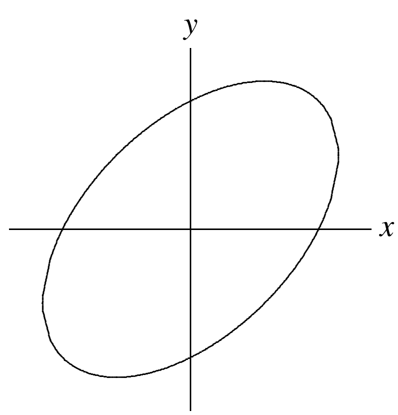


Figure 14.19: $x^2 - xy + y^2 = 2$

There is a similar change of variables formula for triple integrals, though it is a bit more difficult to derive. Suppose we use three substitution functions, $x = f(u,v,w)$, $y = g(u,v,w)$, and $z = h(u,v,w)$. The Jacobian determinant is now

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} f_u & g_u & h_u \\ f_v & g_v & h_v \\ f_w & g_w & h_w \end{vmatrix}.$$

Then the integral is transformed in a similar fashion:

$$\int\int\int_R F(x,y,z)\,dV = \int\int\int_S F(f(u,v,w), g(u,v,w), h(u,v,w)) \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| du\,dv\,dw,$$

where of course the region S in uvw space corresponds to the region R in xyz space.

Exercises for 14.7

Exercise 14.7.1 *Complete Example 14.17 by converting to polar coordinates and evaluating the integral.*

Exercise 14.7.2 *Evaluate* $\iint xy\,dx\,dy$ *over the square with corners* $(0,0)$, $(1,1)$, $(2,0)$, *and* $(1,-1)$ *in two ways: directly, and using* $x=(u+v)/2$, $y=(u-v)/2$.

Exercise 14.7.3 *Evaluate* $\iint x^2+y^2\,dx\,dy$ *over the square with corners* $(-1,0)$, $(0,1)$, $(1,0)$, *and* $(0,-1)$ *in two ways: directly, and using* $x=(u+v)/2$, $y=(u-v)/2$.

Exercise 14.7.4 *Evaluate* $\iint (x+y)e^{x-y}\,dx\,dy$ *over the triangle with corners* $(0,0)$, $(-1,1)$, *and* $(1,1)$ *in two ways: directly, and using* $x=(u+v)/2$, $y=(u-v)/2$.

Exercise 14.7.5 *Evaluate* $\iint y(x-y)\,dx\,dy$ *over the parallelogram with corners* $(0,0)$, $(3,3)$, $(7,3)$, *and* $(4,0)$ *in two ways: directly, and using* $x=u+v$, $y=u$.

Exercise 14.7.6 *Evaluate* $\iint \sqrt{x^2+y^2}\,dx\,dy$ *over the triangle with corners* $(0,0)$, $(4,4)$, *and* $(4,0)$ *using* $x=u$, $y=uv$.

Exercise 14.7.7 *Evaluate* $\iint y\sin(xy)\,dx\,dy$ *over the region bounded by* $xy=1$, $xy=4$, $y=1$, *and* $y=4$ *using* $x=u/v$, $y=v$.

Exercise 14.7.8 *Evaluate* $\iint \sin(9x^2+4y^2)\,dA$, *over the region in the first quadrant bounded by the ellipse* $9x^2+4y^2=1$.

Exercise 14.7.9 *Compute the Jacobian for the substitutions* $x=\rho\sin\phi\cos\theta$, $y=\rho\sin\phi\sin\theta$, $z=\rho\cos\phi$.

Exercise 14.7.10 *Evaluate* $\iiint_E dV$ *where* E *is the solid enclosed by the ellipsoid*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,$$

using the transformation $x=au$, $y=bv$, *and* $z=cw$.

15. Vector Functions

15.1 Curves of Vector Functions

We have already seen that a convenient way to describe a line in three dimensions is to provide a vector that "points to" every point on the line as a parameter t varies, like

$$\langle 1,2,3\rangle + t\langle 1,-2,2\rangle = \langle 1+t, 2-2t, 3+2t\rangle.$$

Except that this gives a particularly simple geometric object, there is nothing special about the individual functions of t that make up the coordinates of this vector—any vector with a parameter, like $\langle f(t), g(t), h(t)\rangle$, will describe some curve in three dimensions as t varies through all possible values.

Example 15.1: Describing Curves

Describe the curves $\langle \cos t, \sin t, 0\rangle$, $\langle \cos t, \sin t, t\rangle$, *and* $\langle \cos t, \sin t, 2t\rangle$.

Solution. As t varies, the first two coordinates in all three functions trace out the points on the unit circle, starting with $(1,0)$ when $t=0$ and proceeding counter-clockwise around the circle as t increases. In the first case, the z coordinate is always 0, so this describes precisely the unit circle in the x-y plane. In the second case, the x and y coordinates still describe a circle, but now the z coordinate varies, so that the height of the curve matches the value of t. When $t=\pi$, for example, the resulting vector is $\langle -1, 0, \pi\rangle$. A bit of thought should convince you that the result is a helix. In the third vector, the z coordinate varies twice as fast as the parameter t, so we get a stretched out helix. Both are shown in Figure 15.1. On the left is the first helix, shown for t between 0 and 4π; on the right is the second helix, shown for t between 0 and 2π. Both start and end at the same point, but the first helix takes two full "turns" to get there, because its z coordinate grows more slowly. ♣

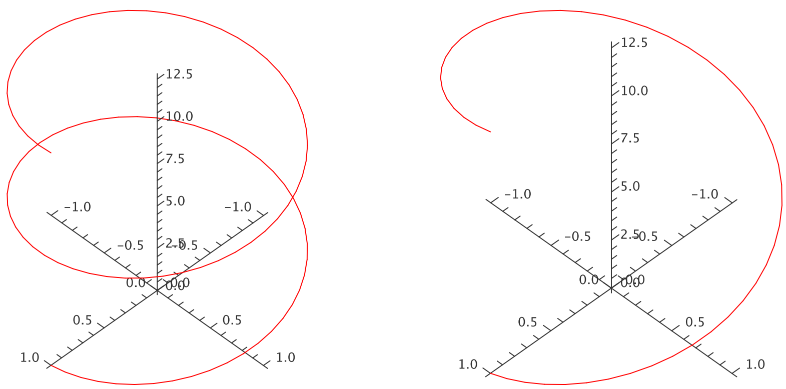


Figure 15.1: Two helixes.

A vector expression of the form $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle$ is called a **vector function**; it is a function from the real numbers $\mathbb{R}$ to the set of all three-dimensional vectors. We can alternately think of it as three separate functions, $x = f(t)$, $y = g(t)$, and $z = h(t)$, that describe points in space. In this case we usually refer to the set of equations as **parametric equations** for the curve, just as for a line. While the parameter t in a vector function might represent any one of a number of physical quantities, or be simply a "pure number", it is often convenient and useful to think of t as representing time. The vector function then tells you where in space a particular object is at any time.

Vector functions can be difficult to understand, that is, difficult to picture. When available, computer software can be very helpful. When working by hand, one useful approach is to consider the "projections" of the curve onto the three standard coordinate planes. We have already done this in part: in Example 15.1 we noted that all three curves project to a circle in the x-y plane, since $\langle \cos t, \sin t\rangle$ is a two dimensional vector function for the unit circle.

Example 15.2: Projections onto a Plane

Graph the projections of $\langle \cos t, \sin t, 2t\rangle$ onto the x-z plane and the y-z plane.

Solution. The two dimensional vector function for the projection onto the x-z plane is $\langle \cos t, 2t\rangle$, or in parametric form, $x = \cos t$, $z = 2t$. By eliminating t we get the equation $x = \cos(z/2)$, the familiar curve shown on the left in Figure 15.2. For the projection onto the y-z plane, we start with the vector function $\langle \sin t, 2t\rangle$, which is the same as $y = \sin t$, $z = 2t$. Eliminating t gives $y = \sin(z/2)$, as shown on the right in Figure 15.2. ♣

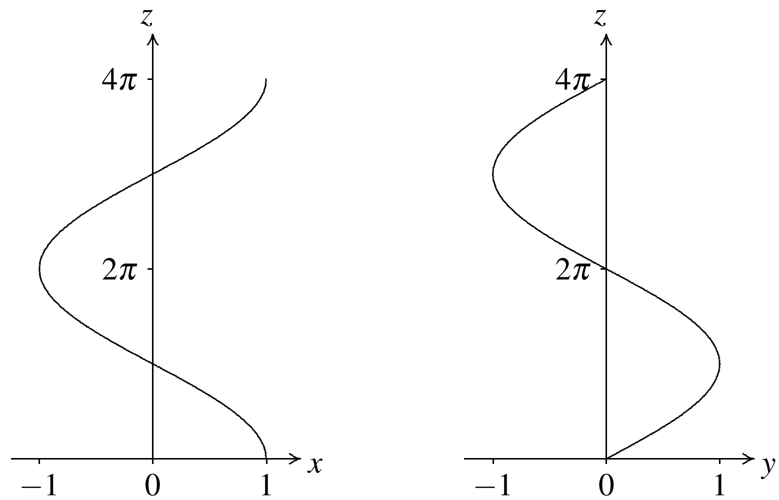


Figure 15.2: The projections of $\langle \cos t, \sin t, 2t \rangle$ onto the x-z and y-z planes.

Exercises for 15.1

Exercise 15.1.1 *Investigate the curve* $\mathbf{r}(t) = \langle \sin t, \cos t, \cos 8t \rangle$.

Exercise 15.1.2 *Investigate the curve* $\mathbf{r}(t) = \langle t \cos t, t \sin t, t \rangle$.

Exercise 15.1.3 *Investigate the curve* $\mathbf{r}(t) = \langle t, t^2, \cos t \rangle$.

Exercise 15.1.4 *Investigate the curve* $\mathbf{r}(t) = \langle \cos(20t)\sqrt{1-t^2}, \sin(20t)\sqrt{1-t^2}, t \rangle$

Exercise 15.1.5 *Find a vector function for the curve of intersection of* $x^2 + y^2 = 9$ *and* $y + z = 2$.

Exercise 15.1.6 *A bug is crawling outward along the spoke of a wheel that lies along a radius of the wheel. The bug is crawling at 1 unit per second and the wheel is rotating at 1 radian per second. Suppose the wheel lies in the y-z plane with center at the origin, and at time* $t = 0$ *the spoke lies along the positive y axis and the bug is at the origin. Find a vector function* $\mathbf{r}(t)$ *for the position of the bug at time* t.

Exercise 15.1.7 *What is the difference between the parametric curves* $f(t) = \langle t, t, t^2 \rangle$, $g(t) = \langle t^2, t^2, t^4 \rangle$, *and* $h(t) = \langle \sin(t), \sin(t), \sin^2(t) \rangle$ *as t runs over all real numbers?*

Exercise 15.1.8 *Plot each of the curves below in 2 dimensions, projected onto each of the three standard planes (the x-y, x-z, and y-z planes).*

(a) $f(t) = \langle t, t^3, t^2 \rangle$, *t ranges over all real numbers*

(b) $f(t) = \langle t^2, t-1, t^2+5 \rangle$ *for* $0 \le t \le 3$

Exercise 15.1.9 *Given points $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$, give parametric equations for the line* segment *connecting A and B. Be sure to give appropriate t values.*

Exercise 15.1.10 *With a parametric plot and a set of t values, we can associate a 'direction'. For example, the curve $\langle \cos t, \sin t\rangle$ is the unit circle traced counterclockwise. How can we amend a set of given parametric equations and t values to get the same curve, only traced backwards?*

15.2 Calculus with Vector Functions

A vector function $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle$ is a function of one variable—that is, there is only one "input" value. What makes vector functions more complicated than the functions $y = f(x)$ that we studied in the first part of this book is of course that the "output" values are now three-dimensional vectors instead of simply numbers. It is natural to wonder if there is a corresponding notion of derivative for vector functions. In the simpler case of a function $y = s(t)$, in which t represents time and $s(t)$ is position on a line, we have seen that the derivative $s'(t)$ represents velocity; we might hope that in a similar way the derivative of a vector function would tell us something about the velocity of an object moving in three dimensions.

One way to approach the question of the derivative for vector functions is to write down an expression that is analogous to the derivative we already understand, and see if we can make sense of it. If we say that what we mean by the limit of a vector is the vector of the individual coordinate limits, this gives us

$$\begin{aligned}
\mathbf{r}'(t) &= \lim_{\Delta t\to 0} \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t)}{\Delta t} \\
&= \lim_{\Delta t\to 0} \frac{\langle f(t+\Delta t) - f(t), g(t+\Delta t) - g(t), h(t+\Delta t) - h(t)\rangle}{\Delta t} \\
&= \lim_{\Delta t\to 0} \langle \frac{f(t+\Delta t) - f(t)}{\Delta t}, \frac{g(t+\Delta t) - g(t)}{\Delta t}, \frac{h(t+\Delta t) - h(t)}{\Delta t}\rangle \\
&= \langle f'(t), g'(t), h'(t)\rangle,
\end{aligned}$$

Starting with a familiar expression for what appears to be a derivative, we find that we can make good computational sense out of it—but what does it actually mean?

We know how to interpret $\mathbf{r}(t+\Delta t)$ and $\mathbf{r}(t)$—they are vectors that point to locations in space; if t is time, we can think of these points as positions of a moving object at times that are Δt apart. We also know what $\Delta\mathbf{r} = \mathbf{r}(t+\Delta t) - \mathbf{r}(t)$ means—it is a vector that points from the head of $\mathbf{r}(t)$ to the head of $\mathbf{r}(t+\Delta t)$, assuming both have their tails at the origin. So when Δt is small, $\Delta\mathbf{r}$ is a tiny vector pointing from one point on the path of the object to a nearby point. As Δt gets close to 0, this vector points in a direction that is closer and closer to the direction in which the object is moving; geometrically, it approaches a vector tangent to the path of the object at a particular point.

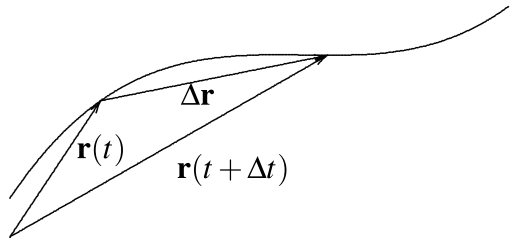


Figure 15.3: Approximating the derivative.

Unfortunately, the vector $\Delta\mathbf{r}$ approaches 0 in length; the vector $\langle 0,0,0\rangle$ is not very informative. By dividing by Δt, when it is small, we effectively keep magnifying the length of $\Delta\mathbf{r}$ so that in the limit it doesn't disappear. Thus the limiting vector $\langle f'(t), g'(t), h'(t)\rangle$ will (usually) be a good, non-zero vector that is tangent to the curve.

What about the length of this vector? It's nice that we've kept it away from zero, but what does it measure, if anything? Consider the length of one of the vectors that approaches the tangent vector:

$$\left|\frac{\mathbf{r}(t+\Delta t)-\mathbf{r}(t)}{\Delta t}\right| = \frac{|\mathbf{r}(t+\Delta t)-\mathbf{r}(t)|}{|\Delta t|}$$

The numerator is the length of the vector that points from one position of the object to a "nearby" position; this length is approximately the distance travelled by the object between times t and $t+\Delta t$. Dividing this distance by the length of time it takes to travel that distance gives the average speed. As Δt approaches zero, this average speed approaches the actual, instantaneous speed of the object at time t.

By performing an "obvious" calculation to get something that looks like the derivative of $\mathbf{r}(t)$, we get precisely what we would want from such a derivative: the vector $\mathbf{r}'(t)$ points in the direction of travel of the object and its length tells us the speed of travel. In the case that t is time, then, we call $\mathbf{v}(t) = \mathbf{r}'(t)$ the velocity vector. Even if t is not time, $\mathbf{r}'(t)$ is useful—it is a vector tangent to the curve.

Example 15.3:

We have seen that $\mathbf{r} = \langle \cos t, \sin t, t\rangle$ is a helix. Compute $\mathbf{r}'$ and $|\mathbf{r}'|$.

Solution. We compute $\mathbf{r}' = \langle -\sin t, \cos t, 1\rangle$, and $|\mathbf{r}'| = \sqrt{\sin^2 t + \cos^2 t + 1} = \sqrt{2}$. Thinking of this as a description of a moving object, its speed is always $\sqrt{2}$; see Figure 15.4. ♣

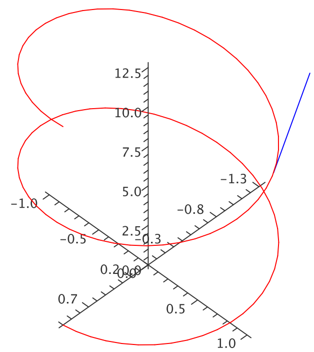


Figure 15.4: A tangent vector on the helix.

Example 15.4:

The velocity vector for $\langle \cos t, \sin t, \cos t \rangle$ *is* $\langle -\sin t, \cos t, -\sin t \rangle$*. As before, the first two coordinates mean that from above this curve looks like a circle. The* z *coordinate is now also periodic, so that as the object moves around the curve its height oscillates up and down. In fact it turns out that the curve is a tilted ellipse, as shown in Figure 15.5.*

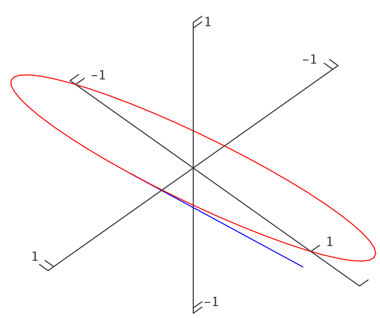


Figure 15.5: The ellipse $\mathbf{r} = \langle \cos t, \sin t, \cos t \rangle$.

Example 15.5:

The velocity vector for $\langle \cos t, \sin t, \cos 2t \rangle$ *is* $\langle -\sin t, \cos t, -2\sin 2t \rangle$*. The* z *coordinate is now oscillating twice as fast as in the previous example, so the graph is not surprising; see Figure 15.6.*

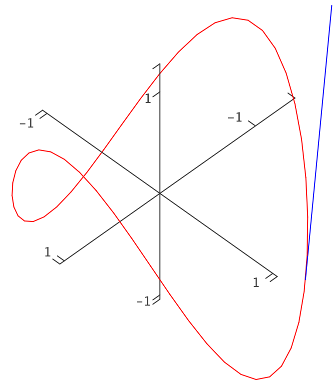


Figure 15.6: $\langle \cos t, \sin t, \cos 2t \rangle$.

Example 15.6:

Find the angle between the curves $\langle t+1, t-2, t^2-4 \rangle$ and $\langle t+2, -t+3, t^2+1 \rangle$ where they meet.

Solution. The angle between two curves at a point is the angle between their tangent vectors—any tangent vectors will do, so we can use the derivatives. We need to find the point of intersection, evaluate the two derivatives there, and finally find the angle between them.

To find the point of intersection, we need to solve the equations

$$\begin{aligned} t+1 &= u+2 \\ t-2 &= -u+3 \\ t^2-4 &= u^2+1 \end{aligned}$$

Solving either of the first two equations for u and substituting in the third gives $t=3$. This together with $u=2$ satisfies all three equations. Thus the two curves meet at $(4,1,5)$, the first curve when $t=3$ and the second curve when $t=2$.

The derivatives are $\langle 1,1,2t \rangle$ and $\langle 1,-1,2t \rangle$; at the intersection point these are $\langle 1,1,6 \rangle$ and $\langle 1,-1,4 \rangle$. The cosine of the angle between them is then

$$\cos\theta = \frac{1-1+24}{\sqrt{38}\sqrt{18}} = \frac{4}{\sqrt{19}},$$

so $\theta = \arccos(4/\sqrt{19}) \approx 0.41$. ♣

The derivatives of vector functions obey some familiar looking rules, which we will occasionally need.

Theorem 15.7: Vector Derivative Properties

Suppose $\mathbf{r}(t)$ and $\mathbf{s}(t)$ are differentiable vector functions, $f(t)$ is a differentiable function, and a is a real number.

1. $\frac{d}{dt}a\mathbf{r}(t) = a\mathbf{r}'(t)$
2. $\frac{d}{dt}(\mathbf{r}(t)+\mathbf{s}(t)) = \mathbf{r}'(t)+\mathbf{s}'(t)$
3. $\frac{d}{dt}f(t)\mathbf{r}(t) = f(t)\mathbf{r}'(t)+f'(t)\mathbf{r}(t)$
4. $\frac{d}{dt}(\mathbf{r}(t)\cdot\mathbf{s}(t)) = \mathbf{r}'(t)\cdot\mathbf{s}(t)+\mathbf{r}(t)\cdot\mathbf{s}'(t)$
5. $\frac{d}{dt}(\mathbf{r}(t)\times\mathbf{s}(t)) = \mathbf{r}'(t)\times\mathbf{s}(t)+\mathbf{r}(t)\times\mathbf{s}'(t)$
6. $\frac{d}{dt}\mathbf{r}(f(t)) = \mathbf{r}'(f(t))f'(t)$

Note that because the cross product is not commutative you must remember to do the three cross products in formula (5.) in the correct order.

When the derivative of a function $f(t)$ is zero, we know that the function has a horizontal tangent line, and may have a local maximum or minimum point. If $\mathbf{r}'(t) = \mathbf{0}$, the geometric interpretation is quite different, though the interpretation in terms of motion is similar. Certainly we know that the object has speed zero at such a point, and it may thus be abruptly changing direction. In three dimensions there are many ways to change direction; geometrically this often means the curve has a cusp or a point, as in the path of a ball that bounces off the floor or a wall.

Example 15.8:

Suppose that $\mathbf{r}(t) = \langle 1+t^3, t^2, 1\rangle$, so $\mathbf{r}'(t) = \langle 3t^2, 2t, 0\rangle$. This is $\mathbf{0}$ at $t = 0$, and there is indeed a cusp at the point $(1,0,1)$, as shown in Figure 15.7.

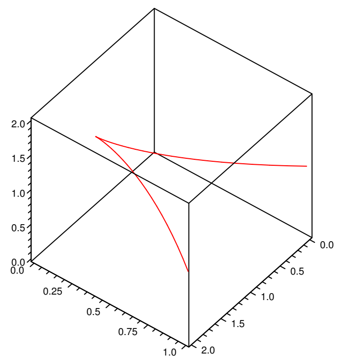


Figure 15.7: $\langle 1+t^3, t^2, 1\rangle$ **has a cusp at** $\langle 1,0,1\rangle$**.**

Sometimes we will be interested in the direction of $\mathbf{r}'$ but not its length. In some cases, we can still work with $\mathbf{r}'$, as when we find the angle between two curves. On other occasions it will be useful to work with a unit vector in the same direction as $\mathbf{r}'$; of course, we can compute such a vector by dividing $\mathbf{r}'$ by its own length. This standard unit tangent vector is usually denoted by $\mathbf{T}$:

$$\mathbf{T} = \frac{\mathbf{r}'}{|\mathbf{r}'|}.$$

In a sense, when we computed the angle between two tangent vectors we have already made use of the unit tangent, since

$$\cos\theta = \frac{\mathbf{r}'\cdot\mathbf{s}'}{|\mathbf{r}'||\mathbf{s}'|} = \frac{\mathbf{r}'}{|\mathbf{r}'|}\cdot\frac{\mathbf{s}'}{|\mathbf{s}'|}$$

Now that we know how to make sense of $\mathbf{r}'$, we immediately know what an antiderivative must be, namely

$$\int \mathbf{r}(t)\,dt = \langle \int f(t)\,dt, \int g(t)\,dt, \int h(t)\,dt\rangle,$$

if $\mathbf{r} = \langle f(t), g(t), h(t)\rangle$. What about definite integrals? Suppose that $\mathbf{v}(t)$ gives the velocity of an object at time t. Then $\mathbf{v}(t)\Delta t$ is a vector that approximates the displacement of the object over the time Δt: $\mathbf{v}(t)\Delta t$ points in the direction of travel, and $|\mathbf{v}(t)\Delta t| = |\mathbf{v}(t)||\Delta t|$ is the speed of the object times Δt, which is approximately the distance travelled. Thus, if we sum many such tiny vectors:

$$\sum_{i=0}^{n-1}\mathbf{v}(t_i)\Delta t$$

we get an approximation to the displacement vector over the time interval $[t_0, t_n]$. If we take the limit we get the exact value of the displacement vector:

$$\lim\sum_{i=0}^{n-1}\mathbf{v}(t_i)\Delta t = \int_{t_0}^{t_n}\mathbf{v}(t)\,dt = \mathbf{r}(t_n) - \mathbf{r}(t_0).$$

Thus, given the velocity vector we can compute the vector function $\mathbf{r}$ giving the location of the object:

$$\mathbf{r}(t) = \mathbf{r}_0 + \int_0^t \mathbf{v}(u)\,du.$$

Example 15.9:

An object moves with velocity vector $\langle \cos t, \sin t, \cos t\rangle$, starting at $(1,1,1)$. Find the function $\mathbf{r}$ giving its location.

Solution.

$$\begin{aligned}
\mathbf{r}(t) &= \langle 1,1,1\rangle + \int_0^t \langle \cos u, \sin u, \cos u\rangle\, du \\
&= \langle 1,1,1\rangle + \langle \sin u, -\cos u, \sin u\rangle\big|_0^t \\
&= \langle 1,1,1\rangle + \langle \sin t, -\cos t, \sin t\rangle - \langle 0,-1,0\rangle \\
&= \langle 1+\sin t, 2-\cos t, 1+\sin t\rangle
\end{aligned}$$

See Figure 15.8. ♣

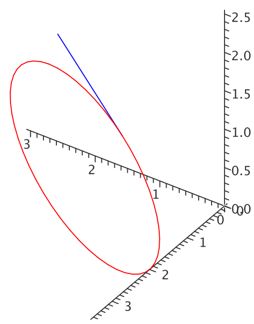


Figure 15.8: Path of the object with its initial velocity vector.

Exercises for 15.2

Exercise 15.2.1 *Find $\mathbf{r}'$ and $\mathbf{T}$ for $\mathbf{r} = \langle t^2, 1, t\rangle$.*

Exercise 15.2.2 *Find $\mathbf{r}'$ and $\mathbf{T}$ for $\mathbf{r} = \langle \cos t, \sin 2t, t^2\rangle$.*

Exercise 15.2.3 *Find $\mathbf{r}'$ and $\mathbf{T}$ for $\mathbf{r} = \langle \cos(e^t), \sin(e^t), \sin t\rangle$.*

Exercise 15.2.4 *Find a vector function for the line tangent to the helix $\langle \cos t, \sin t, t\rangle$ when $t = \pi/4$.*

Exercise 15.2.5 *Find a vector function for the line tangent to $\langle \cos t, \sin t, \cos 4t\rangle$ when $t = \pi/3$.*

Exercise 15.2.6 *Find the cosine of the angle between the curves* $\langle 0, t^2, t\rangle$ *and* $\langle \cos(\pi t/2), \sin(\pi t/2), t\rangle$ *where they intersect.*

Exercise 15.2.7 *Find the cosine of the angle between the curves* $\langle \cos t, -\sin(t)/4, \sin t\rangle$ *and* $\langle \cos t, \sin t, \sin(2t)\rangle$ *where they intersect.*

Exercise 15.2.8 *Suppose that* $|\boldsymbol{r}(t)| = k$*, for some constant k. This means that* $\boldsymbol{r}$ *describes some path on the sphere of radius k with center at the origin. Show that* $\boldsymbol{r}$ *is perpendicular to* $\boldsymbol{r}'$ *at every point. Hint: Use Theorem 15.7, part (d).*

Exercise 15.2.9 *A bug is crawling along the spoke of a wheel that lies along a radius of the wheel. The bug is crawling at 1 unit per second and the wheel is rotating at 1 radian per second. Suppose the wheel lies in the y-z plane with center at the origin, and at time* $t = 0$ *the spoke lies along the positive y axis and the bug is at the origin. Find a vector function* $\boldsymbol{r}(t)$ *for the position of the bug at time t, the velocity vector* $\boldsymbol{r}'(t)$*, the unit tangent* $\boldsymbol{T}(t)$*, and the speed of the bug* $|\boldsymbol{r}'(t)|$*.*

Exercise 15.2.10 *An object moves with velocity vector* $\langle \cos t, \sin t, t\rangle$*, starting at* $\langle 0,0,0\rangle$ *when* $t = 0$*. Find the function* $\boldsymbol{r}$ *giving its location.*

Exercise 15.2.11 *The position function of a particle is given by* $\boldsymbol{r}(t) = \langle t^2, 5t, t^2 - 16t\rangle$*,* $t \geq 0$*. When is the speed of the particle a minimum?*

Exercise 15.2.12 *A particle moves so that its position is given by* $\langle \cos t, \sin t, \cos(6t)\rangle$*. Find the maximum and minimum speeds of the particle.*

Exercise 15.2.13 *An object moves with velocity vector* $\langle t, t^2, \cos t\rangle$*, starting at* $\langle 0,0,0\rangle$ *when* $t = 0$*. Find the function* $\boldsymbol{r}$ *giving its location.*

Exercise 15.2.14 *What is the physical interpretation of the dot product of two vector valued functions? What is the physical interpretation of the cross product of two vector valued functions?*

Exercise 15.2.15 *Show, using the rules of cross products and differentiation, that*

$$\frac{d}{dt}(\boldsymbol{r}(t) \times \boldsymbol{r}'(t)) = \boldsymbol{r}(t) \times \boldsymbol{r}''(t).$$

Exercise 15.2.16 *Determine the point at which* $\boldsymbol{f}(t) = \langle t, t^2, t^3\rangle$ *and* $\boldsymbol{g}(t) = \langle \cos(t), \cos(2t), t+1\rangle$ *intersect, and find the angle between the curves at that point. (Hint: You'll need to set this one up like a line intersection problem, writing one in s and one in t.) If these two functions were the trajectories of two airplanes on the same scale of time, would the planes collide at their point of intersection? Explain.*

Exercise 15.2.17 *Find the equation of the plane perpendicular to the curve* $\boldsymbol{r}(t) = \langle 2\sin(3t), t, 2\cos(3t)\rangle$ *at the point* $(0, \pi, -2)$*.*

Exercise 15.2.18 *Find the equation of the plane perpendicular to* $\langle \cos t, \sin t, \cos(6t)\rangle$ *when* $t = \pi/4$*.*

Exercise 15.2.19 *At what point on the curve $\mathbf{r}(t) = \langle t^3, 3t, t^4\rangle$ is the plane perpendicular to the curve also parallel to the plane $6x + 6y - 8z = 1$?*

Exercise 15.2.20 *Find the equation of the line tangent to $\langle \cos t, \sin t, \cos(6t)\rangle$ when $t = \pi/4$.*

15.3 Arc Length

Sometimes it is useful to compute the length of a curve in space; for example, if the curve represents the path of a moving object, the length of the curve between two points may be the distance travelled by the object between two times.

Recall that if the curve is given by the vector function $\mathbf{r}$ then the vector $\Delta\mathbf{r} = \mathbf{r}(t + \Delta t) - \mathbf{r}(t)$ points from one position on the curve to another, as depicted in Figure 15.3. If the points are close together, the length of $\Delta\mathbf{r}$ is close to the length of the curve between the two points. If we add up the lengths of many such tiny vectors, placed head to tail along a segment of the curve, we get an approximation to the length of the curve over that segment. In the limit, as usual, this sum turns into an integral that computes precisely the length of the curve. First, note that

$$|\Delta\mathbf{r}| = \frac{|\Delta\mathbf{r}|}{\Delta t}\Delta t \approx |\mathbf{r}'(t)|\,\Delta t,$$

when Δt is small. Then the length of the curve between $\mathbf{r}(a)$ and $\mathbf{r}(b)$ is

$$\lim_{n\to\infty}\sum_{i=0}^{n-1}|\Delta\mathbf{r}| = \lim_{n\to\infty}\sum_{i=0}^{n-1}\frac{|\Delta\mathbf{r}|}{\Delta t}\Delta t = \lim_{n\to\infty}\sum_{i=0}^{n-1}|\mathbf{r}'(t)|\,\Delta t = \int_a^b |\mathbf{r}'(t)|\,dt.$$

It is worth stating that this works if between a and b the segment of curve is traced out exactly once.

Recall from Figure 15.1 that the function $\mathbf{r} = \langle \cos t, \sin t, t\rangle$ is a helix.

Example 15.10:

Find the length of one turn of the helix $\mathbf{r} = \langle \cos t, \sin t, t\rangle$ for $\frac{\pi}{2} \le t \le \frac{5\pi}{2}$.

Solution. We compute $\mathbf{r}' = \langle -\sin t, \cos t, 1\rangle$ and $|\mathbf{r}'| = \sqrt{\sin^2 t + \cos^2 t + 1} = \sqrt{2}$, so the length is

$$\int_{\frac{\pi}{2}}^{\frac{5\pi}{2}} \sqrt{2}\,dt = \frac{5\sqrt{2}\pi}{2} - \frac{\sqrt{2}\pi}{2} = 2\sqrt{2}\pi.$$

♣

Example 15.11: Length of a Curve

Suppose $y = \ln x$; what is the length of this curve between $x = 1$ and $x = \sqrt{3}$?

Solution. Although this problem does not appear to involve vectors or three dimensions, we can interpret it in those terms: let $\mathbf{r}(t) = \langle t, \ln t, 0\rangle$. This vector function traces out precisely $y = \ln x$ in the x-y plane. Then $\mathbf{r}'(t) = \langle 1, 1/t, 0\rangle$ and $|\mathbf{r}'(t)| = \sqrt{1+1/t^2}$ and the desired length is

$$\int_1^{\sqrt{3}} \sqrt{1+\frac{1}{t^2}}\,dt = 2-\sqrt{2}+\ln(\sqrt{2}+1)-\frac{1}{2}\ln 3.$$

(This integral is a bit tricky, but requires only methods we have learned.) ♣

Notice that there is nothing special about $y = \ln x$, except that the resulting integral can be computed. In general, given any $y = f(x)$, we can think of this as the vector function $\mathbf{r}(t) = \langle t, f(t), 0\rangle$. Then $\mathbf{r}'(t) = \langle 1, f'(t), 0\rangle$ and $|\mathbf{r}'(t)| = \sqrt{1+(f')^2}$. The length of the curve $y = f(x)$ between a and b is thus

$$\int_a^b \sqrt{1+(f'(x))^2}\,dx.$$

Unfortunately, such integrals are often impossible to do exactly and must be approximated.

One useful application of arc length is the **arc length parameterization**. A vector function $\mathbf{r}(t)$ gives the position of a point in terms of the parameter t, which is often time, but need not be. Suppose s is the distance along the curve from some fixed starting point; if we use s for the variable, we get $\mathbf{r}(s)$, the position in space in terms of distance along the curve. We might still imagine that the curve represents the position of a moving object; now we get the position of the object as a function of how far the object has travelled.

Example 15.12:

Suppose $\mathbf{r}(t) = \langle \cos t, \sin t, 0\rangle$. We know that this curve is a circle of radius 1. While t might represent time, it can also in this case represent the usual angle between the positive x-axis and $\mathbf{r}(t)$. The distance along the circle from $(1,0,0)$ to $(\cos t, \sin t, 0)$ is also t—this is the definition of radian measure. Thus, in this case $s = t$ and $\mathbf{r}(s) = \langle \cos s, \sin s, 0\rangle$.

Example 15.13:

Suppose $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$. We know that this curve is a helix. Find the distance along the helix from $(1,0,0)$ to $(\cos t, \sin t, t)$ and use this to write the arc length parameterization $\hat{\mathbf{r}}(s)$ of $\mathbf{r}(t)$.

Solution. The distance along the helix from $(1,0,0)$ to $(\cos t, \sin t, t)$ is

$$s = \int_0^t |\mathbf{r}'(u)|\,du = \int_0^t \sqrt{\cos^2 u+\sin^2 u+1}\,du = \int_0^t \sqrt{2}\,du = \sqrt{2}t.$$

Thus, the value of t that gets us distance s along the helix is $t = s/\sqrt{2}$, and so the same curve is given by $\hat{\mathbf{r}}(s) = \langle \cos(s/\sqrt{2}), \sin(s/\sqrt{2}), s/\sqrt{2}\rangle$. ♣

In general, if we have a vector function $\mathbf{r}(t)$, to convert it to a vector function in terms of arc length we compute

$$s = \int_a^t |\mathbf{r}'(u)|\,du = f(t),$$

solve $s = f(t)$ for t, getting $t = g(s)$, and substitute this back into $\mathbf{r}(t)$ to get $\hat{\mathbf{r}}(s) = \mathbf{r}(g(s))$.

Suppose that t is time. By the Fundamental Theorem of Calculus, if we start with arc length

$$s(t) = \int_a^t |\mathbf{r}'(u)|\,du$$

and take the derivative, we get

$$s'(t) = |\mathbf{r}'(t)|.$$

Here $s'(t)$ is the rate at which the arc length is changing, and we have seen that $|\mathbf{r}'(t)|$ is the speed of a moving object; these are of course the same.

Suppose that $\mathbf{r}(s)$ is given in terms of arc length; what is $|\mathbf{r}'(s)|$? It is the rate at which arc length is changing *relative to arc length*; it must be 1! In the case of the helix, for example, the arc length parameterization is $\langle \cos(s/\sqrt{2}), \sin(s/\sqrt{2}), s/\sqrt{2}\rangle$, the derivative is $\langle -\sin(s/\sqrt{2})/\sqrt{2}, \cos(s/\sqrt{2})/\sqrt{2}, 1/\sqrt{2}\rangle$, and the length of this is

$$\sqrt{\frac{\sin^2(s/\sqrt{2})}{2} + \frac{\cos^2(s/\sqrt{2})}{2} + \frac{1}{2}} = \sqrt{\frac{1}{2} + \frac{1}{2}} = 1.$$

So in general, $\mathbf{r}'$ is a unit tangent vector.

Exercises for 15.3

Exercise 15.3.1 *Find the length of* $\langle 3\cos t, 2t, 3\sin t\rangle$, $t \in [0, 2\pi]$.

Exercise 15.3.2 *Find the length of* $\langle t^2, 2, t^3\rangle$, $t \in [0, 1]$.

Exercise 15.3.3 *Find the length of* $\langle t^2, \sin t, \cos t\rangle$, $t \in [0, 1]$.

Exercise 15.3.4 *Find the length of the curve* $y = x^{3/2}$, $x \in [1, 9]$.

Exercise 15.3.5 *Set up an integral to compute the length of* $\langle \cos t, \sin t, e^t\rangle$, $t \in [0, 5]$. *(It is tedious but not too difficult to compute this integral.)*

15.4 Curvature

Given a curve $\mathbf{r}(t)$, we would like to be able to measure, at various points, how sharply curved it is. Clearly this is related to how "fast" a tangent vector is changing direction, so a first guess might be that we can measure curvature with $|\mathbf{r}''(t)|$. A little thought shows that this is flawed; if we think of t as time, for example, we could be tracing out the curve more or less quickly as time passes. The second derivative

$|\mathbf{r}''(t)|$ incorporates this notion of time, so it depends not simply on the geometric properties of the curve but on how quickly we move along the curve.

Example 15.14:

Consider $\mathbf{r}(t) = \langle \cos t, \sin t, 0\rangle$ *and* $\mathbf{s}(t) = \langle \cos 2t, \sin 2t, 0\rangle$. *Both of these vector functions represent the unit circle in the x-y plane, but if t is interpreted as time, the second describes an object moving twice as fast as the first. Computing the second derivatives, we find* $|\mathbf{r}''(t)| = 1$, $|\mathbf{s}''(t)| = 4$.

To remove the dependence on time, we use the arc length parameterization. If a curve is given by $\mathbf{r}(s)$, then the first derivative $\mathbf{r}'(s)$ is a unit vector, that is, $\mathbf{r}'(s) = \mathbf{T}(s)$. We now compute the second derivative $\mathbf{r}''(s) = \mathbf{T}'(s)$ and use $|\mathbf{T}'(s)|$ as the "official" measure of **curvature**, usually denoted κ.

Example 15.15:

We have seen that the arc length parameterization of a particular helix is

$$\mathbf{r}(s) = \langle \cos(s/\sqrt{2}), \sin(s/\sqrt{2}), s/\sqrt{2}\rangle$$

Computing the second derivative gives $\mathbf{r}''(s) = \langle -\cos(s/\sqrt{2})/2, -\sin(s/\sqrt{2})/2, 0\rangle$ *with length* $1/2$.

What if we are given a curve as a vector function $\mathbf{r}(t)$, where t is not arc length? We have seen that arc length can be difficult to compute; fortunately, we do not need to convert to the arc length parameterization to compute curvature. Instead, let us imagine that we have done this, so we have found $t = g(s)$ and then formed $\hat{\mathbf{r}}(s) = \mathbf{r}(g(s))$. The first derivative $\hat{\mathbf{r}}'(s)$ is a unit tangent vector, so it is the same as the unit tangent vector $\mathbf{T}(t) = \mathbf{T}(g(s))$. Taking the derivative of this we get

$$\frac{d}{ds}\mathbf{T}(g(s)) = \mathbf{T}'(g(s))g'(s) = \mathbf{T}'(t)\frac{dt}{ds}.$$

The curvature is the length of this vector:

$$\kappa = |\mathbf{T}'(t)||\frac{dt}{ds}| = \frac{|\mathbf{T}'(t)|}{|ds/dt|} = \frac{|\mathbf{T}'(t)|}{|\mathbf{r}'(t)|}.$$

(Recall that we have seen that $ds/dt = |\mathbf{r}'(t)|$.) Thus we can compute the curvature by computing only derivatives with respect to t; we do not need to do the conversion to arc length.

Example 15.16:

Returning to the helix, suppose we start with the parameterization $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$. *Calculate the curvature* κ.

Solution. First $\mathbf{r}'(t) = \langle -\sin t, \cos t, 1\rangle$, $|\mathbf{r}'(t)| = \sqrt{2}$, and $\mathbf{T}(t) = \langle -\sin t, \cos t, 1\rangle/\sqrt{2}$. Then

$$\mathbf{T}'(t) = \langle -\cos t, -\sin t, 0\rangle/\sqrt{2}$$

$$|\mathbf{T}'(t)| = 1/\sqrt{2}$$

Finally, $\kappa = 1/\sqrt{2}/\sqrt{2} = 1/2$, as before. ♣

Example 15.17:

Consider this circle of radius a: $\mathbf{r}(t) = \langle a\cos t, a\sin t, 1\rangle$. Find the curvature, κ.

Solution. First $\mathbf{r}'(t) = \langle -a\sin t, a\cos t, 0\rangle$. Then $|\mathbf{r}'(t)| = \sqrt{a^2\sin^2 t + a^2\cos^2 t} = a$, and $\mathbf{T}(t) = \langle -a\sin t, a\cos t, 0\rangle/a$. Now

$$\mathbf{T}'(t) = \langle -a\cos t, -a\sin t, 0\rangle/a$$

and $|\mathbf{T}'(t)| = 1$. Finally,

$$\kappa = 1/a$$

the curvature of a circle is everywhere the inverse of the radius. It is sometimes useful to think of curvature as describing what circle a curve most resembles at a point. The curvature of the helix in the previous example is $1/2$; this means that a small piece of the helix looks very much like a circle of radius 2, as shown in Figure 15.9. ♣

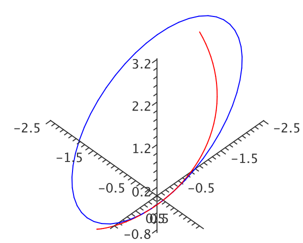


Figure 15.9: A circle with the same curvature as the helix.

Example 15.18:

Consider $\mathbf{r}(t) = \langle \cos t, \sin t, \cos 2t\rangle$, as shown in Figure 15.6. Compute the curvature, κ.

Solution. First $\mathbf{r}'(t) = \langle -\sin t, \cos t, -2\sin(2t)\rangle$ and $|\mathbf{r}'(t)| = \sqrt{1+4\sin^2(2t)}$, so

$$\mathbf{T}(t) = \left\langle \frac{-\sin t}{\sqrt{1+4\sin^2(2t)}}, \frac{\cos t}{\sqrt{1+4\sin^2(2t)}}, \frac{-2\sin 2t}{\sqrt{1+4\sin^2(2t)}} \right\rangle.$$

Computing the derivative of this and then the length of the resulting vector is possible but unpleasant.

Fortunately, there is an alternate formula for the curvature that is often simpler than the one we have:

$$\kappa = \frac{|\mathbf{r}'(t)\times\mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3}.$$

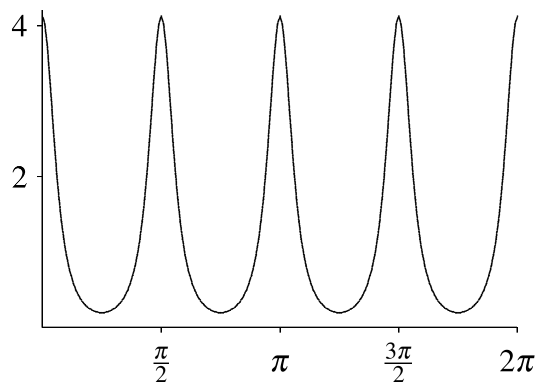
We compute the second derivative $\mathbf{r}''(t) = \langle -\cos t, -\sin t, -4\cos(2t)\rangle$. Then the cross product $\mathbf{r}'(t) \times \mathbf{r}''(t)$ is

$$\langle -4\cos t\cos 2t - 2\sin t\sin 2t, 2\cos t\sin 2t - 4\sin t\cos 2t, 1\rangle.$$

Computing the length of this vector and dividing by $|\mathbf{r}'(t)|^3$ is still a bit tedious. With the aid of a computer we get

$$\kappa = \frac{\sqrt{48\cos^4 t - 48\cos^2 t + 17}}{(-16\cos^4 t + 16\cos^2 t + 1)^{3/2}}.$$

Graphing this we get



Compare this to Figure 15.6. The highest curvature occurs where the curve has its highest and lowest points, and indeed in the picture these appear to be the most sharply curved portions of the curve, while the curve is almost a straight line midway between those points. ♣

Let's see why this alternate formula is correct. Starting with the definition of $\mathbf{T}$, $\mathbf{r}' = |\mathbf{r}'|\mathbf{T}$ so by the product rule $\mathbf{r}'' = |\mathbf{r}'|'\mathbf{T} + |\mathbf{r}'|\mathbf{T}'$. Then by Theorem 12.7 the cross product is

$$\begin{aligned} \mathbf{r}' \times \mathbf{r}'' &= |\mathbf{r}'|\mathbf{T} \times |\mathbf{r}'|'\mathbf{T} + |\mathbf{r}'|\mathbf{T} \times |\mathbf{r}'|\mathbf{T}' \\ &= |\mathbf{r}'||\mathbf{r}'|'(\mathbf{T} \times \mathbf{T}) + |\mathbf{r}'|^2(\mathbf{T} \times \mathbf{T}') \\ &= |\mathbf{r}'|^2(\mathbf{T} \times \mathbf{T}') \end{aligned}$$

because $\mathbf{T} \times \mathbf{T} = \mathbf{0}$, since $\mathbf{T}$ is parallel to itself. Then

$$\begin{aligned} |\mathbf{r}' \times \mathbf{r}''| &= |\mathbf{r}'|^2|\mathbf{T} \times \mathbf{T}'| \\ &= |\mathbf{r}'|^2|\mathbf{T}||\mathbf{T}'|\sin\theta \\ &= |\mathbf{r}'|^2|\mathbf{T}'| \end{aligned}$$

using Exercise 15.2.8 in Section 15.2 to see that $\theta = \pi/2$. Dividing both sides by $|\mathbf{r}'|^3$ then gives the desired formula.

We used the fact here that $\mathbf{T}'$ is perpendicular to $\mathbf{T}$; the vector $\mathbf{N} = \mathbf{T}'/|\mathbf{T}'|$ is thus a unit vector perpendicular to $\mathbf{T}$, called the **unit normal** to the curve. Occasionally of use is the **unit binormal** $\mathbf{B} = \mathbf{T} \times \mathbf{N}$, a unit vector perpendicular to both $\mathbf{T}$ and $\mathbf{N}$.

Exercises for 15.3

Exercise 15.4.1 *Find the curvature of* $\langle t, t^2, t\rangle$.

Exercise 15.4.2 *Find the curvature of* $\langle t, t^2, t^2\rangle$.

Exercise 15.4.3 *Find the curvature of* $\langle t, t^3, t^5\rangle$.

Exercise 15.4.4 *Find the curvature of* $y = x^4$ *at* $(1,1)$.

15.5 Acceleration Vectors

We have already seen that if t is time and an object's location is given by $\mathbf{r}(t)$, then the derivative $\mathbf{r}'(t)$ is the velocity vector $\mathbf{v}(t)$. Just as $\mathbf{v}(t)$ is a vector describing how $\mathbf{r}(t)$ changes, so is $\mathbf{v}'(t)$ a vector describing how $\mathbf{v}(t)$ changes, namely, $\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t)$ is the **acceleration vector**.

Example 15.19:

Suppose $\mathbf{r}(t) = \langle \cos t, \sin t, 1\rangle$. *Then* $\mathbf{r}'(t) = \mathbf{v}(t) = \langle -\sin t, \cos t, 0\rangle$ *and* $\mathbf{v}'(t) = \mathbf{a}(t) = \langle -\cos t, -\sin t, 0\rangle$. *This describes the motion of an object traveling on a circle of radius 1, with constant* z *coordinate 1. The velocity vector is of course tangent to the curve; note that* $\mathbf{a}\cdot\mathbf{v} = 0$, *so* $\mathbf{v}$ *and* $\mathbf{a}$ *are perpendicular. In fact, it is not hard to see that* $\mathbf{a}$ *points from the location of the object to the center of the circular path at* $(0,0,1)$.

Recall that the unit tangent vector is given by $\mathbf{T} = \mathbf{v}/|\mathbf{v}|$, so $\mathbf{v} = |\mathbf{v}|\mathbf{T}$. If we take the derivative of both sides of this equation we get

$$\mathbf{a} = |\mathbf{v}|'\mathbf{T} + |\mathbf{v}|\mathbf{T}'. \tag{15.1}$$

Also recall the definition of the curvature, $\kappa = |\mathbf{T}'|/|\mathbf{v}|$, or $|\mathbf{T}'| = \kappa|\mathbf{v}|$. Finally, recall that we defined the unit normal vector as $\mathbf{N} = \mathbf{T}'/|\mathbf{T}'|$, so $\mathbf{T}' = |\mathbf{T}'|\mathbf{N} = \kappa|\mathbf{v}|\mathbf{N}$. Substituting into Equation 15.1 we get

$$\mathbf{a} = |\mathbf{v}|'\mathbf{T} + \kappa|\mathbf{v}|^2\mathbf{N}. \tag{15.2}$$

The quantity $|\mathbf{v}(t)|$ is the speed of the object, often written as $v(t)$; $|\mathbf{v}(t)|'$ is the rate at which the speed is changing, or the scalar acceleration of the object, $a(t)$. Rewriting Equation 15.2 with these gives us

$$\mathbf{a} = a\mathbf{T} + \kappa v^2\mathbf{N} = a_T\mathbf{T} + a_N\mathbf{N};$$

a_T is the **tangential component of acceleration** and a_N is the **normal component of acceleration**. We have already seen that a_T measures how the speed is changing; if you are riding in a vehicle with large a_T you will feel a force pulling you into your seat. The other component, a_N, measures how sharply your direction is changing *with respect to time*. So it naturally is related to how sharply the path is curved, measured by κ, and also to how fast you are going. Because a_N includes v^2, note that the effect of speed is magnified; doubling your speed around a curve quadruples the value of a_N. You feel the effect of this as a force pushing you toward the outside of the curve, the "centrifugal force."

In practice, if want a_N we would use the formula for κ:

$$a_N = \kappa|\mathbf{v}|^2 = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3}|\mathbf{r}'|^2 = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|} = \frac{|\mathbf{v} \times \mathbf{a}|}{|\mathbf{v}|}.$$

To compute a_T we can project $\mathbf{a}$ onto $\mathbf{v}$:

$$a_T = \frac{\mathbf{v} \cdot \mathbf{a}}{|\mathbf{v}|} = \frac{\mathbf{r}' \cdot \mathbf{r}''}{|\mathbf{r}'|}.$$

Example 15.20:

Suppose $\mathbf{r} = \langle t, t^2, t^3 \rangle$. *Compute* $\mathbf{v}$, $\mathbf{a}$, a_T, *and* a_N.

Solution. Taking derivatives we get $\mathbf{v} = \langle 1, 2t, 3t^2 \rangle$ and $\mathbf{a} = \langle 0, 2, 6t \rangle$. Then

$$a_T = \frac{4t + 18t^3}{\sqrt{1 + 4t^2 + 9t^4}} \quad \text{and} \quad a_N = \frac{\sqrt{4 + 36t^2 + 36t^4}}{\sqrt{1 + 4t^2 + 9t^4}}.$$

♣

Exercises for 15.5

Exercise 15.5.1 *Let* $\boldsymbol{r} = \langle t^5, 2t^2, t \rangle$. *Compute* $\boldsymbol{v}$, $\boldsymbol{a}$, a_T, *and* a_N.

Exercise 15.5.2 *Let* $\boldsymbol{r} = \langle \cos t, \sin t, t^2 \rangle$. *Compute* $\boldsymbol{v}$, $\boldsymbol{a}$, a_T, *and* a_N.

Exercise 15.5.3 *Let* $\boldsymbol{r} = \langle \cos t, \sin t, e^t \rangle$. *Compute* $\boldsymbol{v}$, $\boldsymbol{a}$, a_T, *and* a_N.

Exercise 15.5.4 *Let* $\boldsymbol{r} = \langle e^t, \sin t, e^t \rangle$. *Compute* $\boldsymbol{v}$, $\boldsymbol{a}$, a_T, *and* a_N.

Exercise 15.5.5 *Suppose an object moves so that its acceleration is given by* $\boldsymbol{a} = \langle -3\cos t, -2\sin t, 0 \rangle$. *At time* $t = 0$ *the object is at* $(3,0,0)$ *and its velocity vector is* $\langle 0,2,0 \rangle$. *Find* $\boldsymbol{v}(t)$ *and* $\boldsymbol{r}(t)$ *for the object.*

Exercise 15.5.6 *Suppose an object moves so that its acceleration is given by* $\boldsymbol{a} = \langle -3\cos t, -2\sin t, 0 \rangle$. *At time* $t = 0$ *the object is at* $(3,0,0)$ *and its velocity vector is* $\langle 0,2.1,0 \rangle$. *Find* $\boldsymbol{v}(t)$ *and* $\boldsymbol{r}(t)$ *for the object.*

Exercise 15.5.7 *Suppose an object moves so that its acceleration is given by* $\boldsymbol{a} = \langle -3\cos t, -2\sin t, 0 \rangle$. *At time* $t = 0$ *the object is at* $(3,0,0)$ *and its velocity vector is* $\langle 0,2,1 \rangle$. *Find* $\boldsymbol{v}(t)$ *and* $\boldsymbol{r}(t)$ *for the object.*

Exercise 15.5.8 *Suppose an object moves so that its acceleration is given by* $\boldsymbol{a} = \langle -3\cos t, -2\sin t, 0\rangle$. *At time* $t = 0$ *the object is at* $(3,0,0)$ *and its velocity vector is* $\langle 0, 2.1, 1\rangle$. *Find* $\boldsymbol{v}(t)$ *and* $\boldsymbol{r}(t)$ *for the object.*

Exercise 15.5.9 *Describe a situation in which the normal component of acceleration is 0 and the tangential component of acceleration is non-zero. Is it possible for the tangential component of acceleration to be 0 while the normal component of acceleration is non-zero? Explain. Finally, is it possible for an object to move (not be stationary) so that both the tangential and normal components of acceleration are 0? Explain.*

16. Vector Calculus

16.1 Vector Fields

This chapter is concerned with applying calculus in the context of **vector fields**. A two-dimensional vector field is a function f that maps each point (x,y) in $\mathbb{R}^2$ to a two-dimensional vector $\langle u,v\rangle$, and similarly a three-dimensional vector field maps (x,y,z) to $\langle u,v,w\rangle$. Since a vector has no position, we typically indicate a vector field in graphical form by placing the vector $f(x,y)$ with its tail at (x,y). Figure 16.1 shows a representation of the vector field $f(x,y) = \langle x/\sqrt{x^2+y^2+4}, -y/\sqrt{x^2+y^2+4}\rangle$. For such a graph to be readable, the vectors must be fairly short, which is accomplished by using a different scale for the vectors than for the axes. Such graphs are thus useful for understanding the sizes of the vectors relative to each other but not their absolute size.

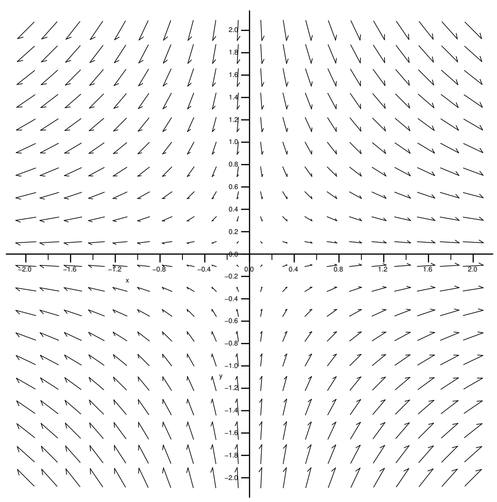


Figure 16.1: A vector field.

Vector fields have many important applications, as they can be used to represent many physical quantities: the vector at a point may represent the strength of some force (gravity, electricity, magnetism) or a velocity (wind speed or the velocity of some other fluid).

We have already seen a particularly important kind of vector field—the gradient. Given a function $f(x,y)$, recall that the gradient is $\mathbf{f} = \nabla f = \langle f_x(x,y), f_y(x,y)\rangle$, a vector that depends on (is a function of) x and y. We usually picture the gradient vector with its tail at (x,y), pointing in the direction of maximum increase. Vector fields that are gradients have some particularly nice properties, as we will see. An important example is

$$\mathbf{f} = \left\langle \frac{-x}{(x^2+y^2+z^2)^{3/2}}, \frac{-y}{(x^2+y^2+z^2)^{3/2}}, \frac{-z}{(x^2+y^2+z^2)^{3/2}} \right\rangle,$$

which points from the point (x,y,z) toward the origin and has length

$$\frac{\sqrt{x^2+y^2+z^2}}{(x^2+y^2+z^2)^{3/2}} = \frac{1}{(\sqrt{x^2+y^2+z^2})^2},$$

which is the reciprocal of the square of the distance from (x,y,z) to the origin—in other words, **f** is an "inverse square law". The vector **f** is a gradient:

$$\mathbf{f} = \nabla \frac{1}{\sqrt{x^2+y^2+z^2}} \tag{16.1}$$

which turns out to be extremely useful.

Exercises for 16.1

Exercise 16.1.1 *Investigate the vector field* $\langle x,y\rangle$

Exercise 16.1.2 *Investigate the vector field* $\langle -x,-y\rangle$

Exercise 16.1.3 *Investigate the vector field* $\langle x,-y\rangle$

Exercise 16.1.4 *Investigate the vector field* $\langle \sin x,\cos y\rangle$

Exercise 16.1.5 *Investigate the vector field* $\langle x+1,x+3\rangle$

Exercise 16.1.6 *Verify Equation 16.1.*

16.2 Divergence and Curl

Divergence and curl are two measurements of vector fields that are very useful in a variety of applications. Both are most easily understood by thinking of the vector field as representing a flow of a liquid or gas; that is, each vector in the vector field should be interpreted as a velocity vector. Roughly speaking, divergence measures the tendency of the fluid to collect or disperse at a point, and curl measures the tendency of the fluid to swirl around the point. Divergence is a scalar, that is, a single number, while curl is itself a vector. The magnitude of the curl measures how much the fluid is swirling, the direction indicates the axis around which it tends to swirl. These ideas are somewhat subtle in practice, and are beyond the scope of this course.

Recall that if f is a function, the gradient of f is given by

$$\nabla f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle.$$

A useful mnemonic for this (and for the divergence and curl, as it turns out) is to let

$$\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle,$$

that is, we pretend that ∇ is a vector with rather odd looking entries. Recalling that $\langle u, v, w\rangle a = \langle ua, va, wa\rangle$, we can then think of the gradient as

$$\nabla f = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle,$$

that is, we simply multiply the f into the vector.

The divergence and curl can now be defined in terms of this same vector ∇ by using the cross product and dot product.

Divergence

The divergence of a vector field $\mathbf{f} = \langle f_1, f_2, f_3\rangle$ is

$$\nabla \cdot \mathbf{f} = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle \cdot \langle f_1, f_2, f_3\rangle = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}.$$

Curl

The curl of $\mathbf{f}$ is

$$\nabla \times \mathbf{f} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_1 & f_2 & f_3 \end{vmatrix} = \left\langle \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}, \frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x}, \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right\rangle.$$

Here are two simple but useful facts about divergence and curl.

Theorem 16.1: Divergence of Curl is Zero

$\nabla \cdot (\nabla \times \boldsymbol{f}) = 0.$

In words, this says that the divergence of the curl is zero.

Theorem 16.2: Curl of Gradient is Zero

$\nabla \times (\nabla f) = \boldsymbol{0}.$

That is, the curl of a gradient is the zero vector. Recalling that gradients are conservative vector fields, this says that the curl of a conservative vector field is the zero vector. Under suitable conditions, it is also true that if the curl of $\mathbf{f}$ is $\mathbf{0}$ then $\mathbf{f}$ is conservative. (Note that this is exactly the same test that we discussed at the end of Section 16.3.3.)

Example 16.3:

Let $\mathbf{f} = \langle 2e^x - y, -x, e^z\rangle$. Show that $\mathbf{f}$ is conservative and find the function f such that $\mathbf{f} = \langle f_x, f_y, f_z\rangle$.

Solution. If $\mathbf{f} = \langle 2e^x - y, -x, e^z\rangle$ then $\nabla\times\mathbf{f} = \langle 0, 0, (-1)-(-1)\rangle = \mathbf{0}$. Thus, $\mathbf{f}$ is conservative, and we can exhibit this directly by finding the corresponding f.

Since $f_x = 2e^x - y$, $f = 2e^x - xy + g(y,z)$. Since $f_y = -x$, it must be that $g_y = 0$, so $g(y,z) = C + h(z)$. Thus $f = 2e^x - xy + C + h(z)$ and

$$f_z = h'(z) = e^z,$$

so $h(z) = e^z$. This leaves $f = 2e^x - xy + e^z + C$. ♣

Exercises for 16.2

Exercise 16.2.1 *Let $\boldsymbol{f} = \langle xy, -xy\rangle$ and let D be given by $0\le x\le 1$, $0\le y\le 1$. Compute $\int_{\partial D}\boldsymbol{f}\cdot d\boldsymbol{r}$ and $\int_{\partial D}\boldsymbol{f}\cdot\boldsymbol{N}\,ds$.*

Exercise 16.2.2 *Let $\boldsymbol{f} = \langle ax^2, by^2\rangle$ and let D be given by $0\le x\le 1$, $0\le y\le 1$. Compute $\int_{\partial D}\boldsymbol{f}\cdot d\boldsymbol{r}$ and $\int_{\partial D}\boldsymbol{f}\cdot\boldsymbol{N}\,ds$.*

Exercise 16.2.3 *Let $\boldsymbol{f} = \langle ay^2, bx^2\rangle$ and let D be given by $0\le x\le 1$, $0\le y\le x$. Compute $\int_{\partial D}\boldsymbol{f}\cdot d\boldsymbol{r}$ and $\int_{\partial D}\boldsymbol{f}\cdot\boldsymbol{N}\,ds$.*

Exercise 16.2.4 *Let $\boldsymbol{f} = \langle \sin x\cos y, \cos x\sin y\rangle$ and let D be given by $0\le x\le \pi/2$, $0\le y\le x$. Compute $\int_{\partial D}\boldsymbol{f}\cdot d\boldsymbol{r}$ and $\int_{\partial D}\boldsymbol{f}\cdot\boldsymbol{N}\,ds$.*

Exercise 16.2.5 *Let $\boldsymbol{f} = \langle y, -x\rangle$ and let D be given by $x^2+y^2\le 1$. Compute $\int_{\partial D}\boldsymbol{f}\cdot d\boldsymbol{r}$ and $\int_{\partial D}\boldsymbol{f}\cdot\boldsymbol{N}\,ds$.*

Exercise 16.2.6 *Let $\boldsymbol{f} = \langle x, y\rangle$ and let D be given by $x^2+y^2\le 1$. Compute $\int_{\partial D}\boldsymbol{f}\cdot d\boldsymbol{r}$ and $\int_{\partial D}\boldsymbol{f}\cdot\boldsymbol{N}\,ds$.*

Exercise 16.2.7 *Prove Theorem 16.1.*

Exercise 16.2.8 *Prove Theorem 16.2.*

Exercise 16.2.9 *If $\nabla\cdot\boldsymbol{f} = 0$, $\boldsymbol{f}$ is said to be* **incompressible**. *Show that any vector field of the form $\boldsymbol{f}(x,y,z) = \langle f(y,z), g(x,z), h(x,y)\rangle$ is incompressible. Give a non-trivial example.*

16.3 Line Integrals

We have so far integrated "over" intervals, areas, and volumes with single, double, and triple integrals. We now investigate integration over or "along" a curve—"line integrals" are really "curve integrals".

16.3.1. Line Integrals of a Function

As with other integrals, a geometric example may be easiest to understand. Consider the function $f = x + y$ and the parabola $y = x^2$ in the x-y plane, for $0 \le x \le 2$. Imagine that we extend the parabola up to the surface f, to form a curved wall or curtain, as in Figure 16.2. What is the area of the surface thus formed? We already know one way to compute surface area, but here we take a different approach that is more useful for the problems to come.

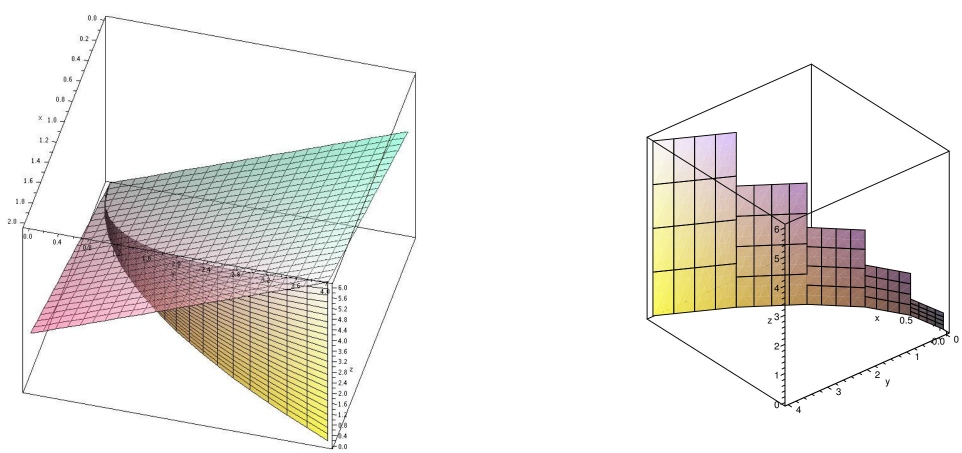


Figure 16.2: Approximating the area under a curve.

As usual, we start by thinking about how to approximate the area. We pick some points along the part of the parabola we're interested in, and connect adjacent points by straight lines; when the points are close together, the length of each line segment will be close to the length along the parabola. Using each line segment as the base of a rectangle, we choose the height to be the height of the surface f above the line segment. If we add up the areas of these rectangles, we get an approximation to the desired area, and in the limit this sum turns into an integral.

Typically the curve is in vector form, or can easily be put in vector form. Then as we have seen in Section 15.3, the length of one of the straight line segments in the approximation is close to $ds = |\mathbf{v}'|\,dt$. In this example we have $\mathbf{v}(t) = \langle t, t^2 \rangle$, and $ds = |\mathbf{v}'|\,dt = \sqrt{1+4t^2}\,dt$. Therefore the integral is

$$\int_0^2 f(t,t^2)\sqrt{1+4t^2}\,dt = \int_0^2 (t+t^2)\sqrt{1+4t^2}\,dt = \frac{167}{48}\sqrt{17} - \frac{1}{12} - \frac{1}{64}\ln(4+\sqrt{17}).$$

This integral of a function along a curve C is often written in abbreviated form as

$$\int_C f(x,y)\,ds.$$

Definition 16.4: Line Integral of a Function

Let $f(x,y,z)$ be a function defined along a curve C. If the curve C is parametrized by $\mathbf{r}(t)$ with $t \in [a,b]$, then the integral of f along C is given by

$$\int_C f(x,y)\,ds = \int_a^b f(\mathbf{r}(t)) \left|\frac{d\mathbf{r}}{dt}\right| dt$$

Example 16.5: Compute Line Segment

Compute $\int_C ye^x\,ds$ where C is the line segment from $(1,2)$ to $(4,7)$.

Solution. We write the line segment as a vector function: $\mathbf{v} = \langle 1,2\rangle + t\langle 3,5\rangle$, $0 \le t \le 1$, or in parametric form $x = 1+3t$, $y = 2+5t$. Then

$$\int_C ye^x\,ds = \int_0^1 (2+5t)e^{1+3t}\sqrt{3^2+5^2}\,dt = \frac{16}{9}\sqrt{34}e^4 - \frac{1}{9}\sqrt{34}\,e.$$

♣

All of these ideas extend to three dimensions in the obvious way.

Example 16.6: Compute Line Segment

Compute $\int_C x^2z\,ds$ where C is the line segment from $(0,6,-1)$ to $(4,1,5)$.

Solution. We write the line segment as a vector function: $\mathbf{v} = \langle 0,6,-1\rangle + t\langle 4,-5,6\rangle$, $0 \le t \le 1$, or in parametric form $x = 4t$, $y = 6-5t$, $z = -1+6t$. Then

$$\int_C x^2z\,ds = \int_0^1 (4t)^2(-1+6t)\sqrt{16+25+36}\,dt = 16\sqrt{77}\int_0^1 -t^2+6t^3\,dt = \frac{56}{3}\sqrt{77}.$$

♣

16.3.2. Line Integrals of a Vector Field

Now we turn to a perhaps more interesting example. Recall that in the simplest case, the work done by a force on an object is equal to the magnitude of the force times the distance the object moves; this assumes that the force is constant and in the direction of motion. We have already dealt with examples in which the force is not constant; now we are prepared to examine what happens when the force is not parallel to the direction of motion.

We have already examined the idea of components of force, in Example 12.5: the component of a force $\mathbf{F}$ in the direction of a vector $\mathbf{v}$ is

$$\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v},$$

the projection of $\mathbf{F}$ onto $\mathbf{v}$. The length of this vector, that is, the magnitude of the force in the direction of $\mathbf{v}$, is

$$\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|},$$

the scalar projection of $\mathbf{F}$ onto $\mathbf{v}$. If an object moves subject to this (constant) force, in the direction of $\mathbf{v}$, over a distance equal to the length of $\mathbf{v}$, the work done is

$$\frac{\mathbf{F}\cdot\mathbf{v}}{|\mathbf{v}|}|\mathbf{v}| = \mathbf{F}\cdot\mathbf{v}.$$

Thus, work in the vector setting is still "force times distance", except that "times" means "dot product".

If the force varies from point to point, it is represented by a vector field $\mathbf{F}$ (note we use a capital F here to denote force). the displacement vector $\mathbf{v}$ may also change, as an object may follow a curving path in two or three dimensions. Suppose that the path of an object is given by a vector function $\mathbf{r}(t)$; at any point along the path, the (small) tangent vector $\mathbf{r}'\,\Delta t$ gives an approximation to its motion over a short time Δt, so the work done during that time is approximately $\mathbf{F}\cdot\mathbf{r}'\,\Delta t$; the total work over some time period is then

$$\int_{t_0}^{t_1} \mathbf{F}\cdot\mathbf{r}'\,dt.$$

It is useful to rewrite this in various ways at different times. We start with

$$\int_{t_0}^{t_1} \mathbf{F}\cdot\mathbf{r}'\,dt = \int_C \mathbf{F}\cdot d\mathbf{r},$$

abbreviating $\mathbf{r}'\,dt$ by $d\mathbf{r}$. Or we can write

$$\int_{t_0}^{t_1} \mathbf{F}\cdot\mathbf{r}'\,dt = \int_{t_0}^{t_1} \mathbf{F}\cdot\frac{\mathbf{r}'}{|\mathbf{r}'|}|\mathbf{r}'|\,dt = \int_{t_0}^{t_1} \mathbf{F}\cdot\mathbf{T}\,|\mathbf{r}'|\,dt = \int_C \mathbf{F}\cdot\mathbf{T}\,ds,$$

using the unit tangent vector $\mathbf{T}$, abbreviating $|\mathbf{r}'|\,dt$ as ds, and indicating the path of the object by C. In other words, work is computed using a particular line integral of the form we have considered. Alternately, we sometimes write

$$\begin{aligned}\int_C \mathbf{F}\cdot\mathbf{r}'\,dt &= \int_C \langle f_1, f_2, f_3\rangle\cdot\langle x', y', z'\rangle\,dt = \int_C \left(f_1\frac{dx}{dt} + f_2\frac{dy}{dt} + f_3\frac{dz}{dt}\right)dt \\ &= \int_C f_1\,dx + f_2\,dy + f_3\,dz = \int_C f_1\,dx + \int_C f_2\,dy + \int_C f_3\,dz,\end{aligned}$$

and similarly for two dimensions, leaving out references to z.

We define this formally.

Definition 16.7: Line Integral of a Vector Field

Let $\boldsymbol{F}$ be a vector field and C a curve with parametrization $\mathbf{r}(t)$. Then the integral of $\boldsymbol{F}$ over C is given by

$$\int_C \boldsymbol{F}\cdot d\mathbf{r} = \int_C \boldsymbol{F}\cdot\mathbf{T}ds = \int_{t_0}^{t_1} \boldsymbol{F}\cdot\mathbf{r}'\,dt$$

Example 16.8: Work Done

Suppose an object moves from $(-1,1)$ to $(2,4)$ along the path $\mathbf{r}(t) = \langle t, t^2 \rangle$, subject to the force $\boldsymbol{F} = \langle x\sin y, y \rangle$. Find the work done.

Solution. We can write the force in terms of t as $\langle t\sin(t^2), t^2 \rangle$, and compute $\mathbf{r}'(t) = \langle 1, 2t \rangle$, and then the work is

$$\int_{-1}^{2} \langle t\sin(t^2), t^2 \rangle \cdot \langle 1, 2t \rangle \, dt = \int_{-1}^{2} t\sin(t^2) + 2t^3 \, dt = \frac{\cos(1) - \cos(4)}{2} + \frac{15}{2}.$$

Alternately, we might write

$$\int_C x\sin y \, dx + \int_C y \, dy = \int_{-1}^{2} x\sin(x^2) \, dx + \int_1^4 y \, dy = -\frac{\cos(4)}{2} + \frac{\cos(1)}{2} + \frac{16}{2} - \frac{1}{2}$$

getting the same answer. ♣

Exercises for 16.3.2

Exercise 16.3.1 *Compute $\int_C xy^2 \, ds$ along the line segment from $(1,2,0)$ to $(2,1,3)$.*

Exercise 16.3.2 *Compute $\int_C \sin x \, ds$ along the line segment from $(-1,2,1)$ to $(1,2,5)$.*

Exercise 16.3.3 *Compute $\int_C z\cos(xy) \, ds$ along the line segment from $(1,0,1)$ to $(2,2,3)$.*

Exercise 16.3.4 *Compute $\int_C xe^y \, dx + x^2 y \, dy$ along the line segment $y = 3$, $0 \le x \le 2$.*

Exercise 16.3.5 *Compute $\int_C xe^y \, dx + x^2 y \, dy$ along the line segment $x = 4$, $0 \le y \le 4$.*

Exercise 16.3.6 *Compute $\int_C xe^y \, dx + x^2 y \, dy$ along the curve $x = 3t$, $y = t^2$, $0 \le t \le 1$.*

Exercise 16.3.7 *Compute $\int_C xe^y \, dx + x^2 y \, dy$ along the curve $\langle e^t, e^t \rangle$, $-1 \le t \le 1$.*

Exercise 16.3.8 *Compute $\int_C \langle \cos x, \sin y \rangle \cdot d\boldsymbol{r}$ along the curve $\langle t, t \rangle$, $0 \le t \le 1$.*

Exercise 16.3.9 *Compute $\int_C \langle 1/xy, 1/(x+y) \rangle \cdot d\boldsymbol{r}$ along the path from $(1,1)$ to $(3,1)$ to $(3,6)$ using straight line segments.*

Exercise 16.3.10 *Compute* $\int_C \langle 1/xy, 1/(x+y)\rangle \cdot d\mathbf{r}$ *along the curve* $\langle 2t, 5t\rangle$, $1 \le t \le 4$.

Exercise 16.3.11 *Compute* $\int_C \langle 1/xy, 1/(x+y)\rangle \cdot d\mathbf{r}$ *along the curve* $\langle t, t^2\rangle$, $1 \le t \le 4$.

Exercise 16.3.12 *Compute* $\int_C yz\,dx + xz\,dy + xy\,dz$ *along the curve* $\langle t, t^2, t^3\rangle$, $0 \le t \le 1$.

Exercise 16.3.13 *Compute* $\int_C yz\,dx + xz\,dy + xy\,dz$ *along the curve* $\langle \cos t, \sin t, \tan t\rangle$, $0 \le t \le \pi$.

Exercise 16.3.14 *An object moves from* $(1,1)$ *to* $(4,8)$ *along the path* $\mathbf{r}(t) = \langle t^2, t^3\rangle$, *subject to the force* $\mathbf{F} = \langle x^2, \sin y\rangle$. *Find the work done.*

Exercise 16.3.15 *An object moves along the line segment from* $(1,1)$ *to* $(2,5)$, *subject to the force* $\mathbf{F} = \langle x/(x^2+y^2), y/(x^2+y^2)\rangle$. *Find the work done.*

Exercise 16.3.16 *An object moves along the parabola* $\mathbf{r}(t) = \langle t, t^2\rangle$, $0 \le t \le 1$, *subject to the force* $\mathbf{F} = \langle 1/(y+1), -1/(x+1)\rangle$. *Find the work done.*

Exercise 16.3.17 *An object moves along the line segment from* $(0,0,0)$ *to* $(3,6,10)$, *subject to the force* $\mathbf{F} = \langle x^2, y^2, z^2\rangle$. *Find the work done.*

Exercise 16.3.18 *An object moves along the curve* $\mathbf{r}(t) = \langle \sqrt{t}, 1/\sqrt{t}, t\rangle$ $1 \le t \le 4$, *subject to the force* $\mathbf{F} = \langle y, z, x\rangle$. *Find the work done.*

Exercise 16.3.19 *An object moves from* $(1,1,1)$ *to* $(2,4,8)$ *along the path* $\mathbf{r}(t) = \langle t, t^2, t^3\rangle$, *subject to the force* $\mathbf{F} = \langle \sin x, \sin y, \sin z\rangle$. *Find the work done.*

Exercise 16.3.20 *An object moves from* $(1,0,0)$ *to* $(-1,0,\pi)$ *along the path* $\mathbf{r}(t) = \langle \cos t, \sin t, t\rangle$, *subject to the force* $\mathbf{F} = \langle y^2, y^2, xz\rangle$. *Find the work done.*

Exercise 16.3.21 *Give an example of a non-trivial force field* $\mathbf{F}$ *and non-trivial path* $\mathbf{r}(t)$ *for which the total work done moving along the path is zero.*

16.3.3. The Fundamental Theorem of Line Integrals

One way to write the Fundamental Theorem of Calculus (Theorem 6.13) is:

$$\int_a^b f'(x)\,dx = f(b) - f(a).$$

That is, to compute the integral of a derivative f' we need only compute the values of f at the endpoints. Something similar is true for line integrals of a certain form.

Theorem 16.9: Fundamental Theorem of Line Integrals

Suppose a curve C is given by the vector function $\mathbf{r}(t)$, with $\mathbf{a} = \mathbf{r}(a)$ and $\mathbf{b} = \mathbf{r}(b)$. Then

$$\int_C \nabla f \cdot d\mathbf{r} = f(\mathbf{b}) - f(\mathbf{a}),$$

provided that $\mathbf{r}$ is sufficiently nice.

Proof. We write $\mathbf{r} = \langle x(t), y(t), z(t)\rangle$, so that $\mathbf{r}' = \langle x'(t), y'(t), z'(t)\rangle$. Also, we know that $\nabla f = \langle f_x, f_y, f_z\rangle$. Then

$$\int_C \nabla f \cdot d\mathbf{r} = \int_a^b \langle f_x, f_y, f_z\rangle \cdot \langle x'(t), y'(t), z'(t)\rangle\, dt = \int_a^b f_x x' + f_y y' + f_z z'\, dt.$$

By the chain rule (see Section 13.4) $f_x x' + f_y y' + f_z z' = df/dt$, where f in this context means $f(x(t), y(t), z(t))$, a function of t. In other words, all we have is

$$\int_a^b f'(t)\, dt = f(b) - f(a).$$

In this context, $f(a) = f(x(a), y(a), z(a))$. Since $\mathbf{a} = \mathbf{r}(a) = \langle x(a), y(a), z(a)\rangle$, we can write $f(a) = f(\mathbf{a})$—this is a bit of a cheat, since we are simultaneously using f to mean $f(t)$ and $f(x,y,z)$, and since $f(x(a), y(a), z(a))$ is not technically the same as $f(\langle x(a), y(a), z(a)\rangle)$, but the concepts are clear and the different uses are compatible. Doing the same for b, we get

$$\int_C \nabla f \cdot d\mathbf{r} = \int_a^b f'(t)\, dt = f(b) - f(a) = f(\mathbf{b}) - f(\mathbf{a}).$$

♣

This theorem, like the Fundamental Theorem of Calculus, says roughly that if we integrate a "derivative-like function" (f' or ∇f) the result depends only on the values of the original function (f) at the endpoints.

If a vector field $\mathbf{f}$ is the gradient of a function, $\mathbf{f} = \nabla f$, we say that $\mathbf{f}$ is a **conservative vector field**. If $\mathbf{F}$ is a conservative force field, then the integral for work, $\int_C \mathbf{F} \cdot d\mathbf{r}$, is in the form required by the Fundamental Theorem of Line Integrals. This means that in a conservative force field, the amount of work required to move an object from point $\mathbf{a}$ to point $\mathbf{b}$ depends only on those points, not on the path taken between them.

Example 16.10: Work Done

An object moves in the force field

$$\mathbf{F} = \left\langle \frac{-x}{(x^2+y^2+z^2)^{3/2}}, \frac{-y}{(x^2+y^2+z^2)^{3/2}}, \frac{-z}{(x^2+y^2+z^2)^{3/2}} \right\rangle,$$

along the curve $\mathbf{r} = \langle 1+t, t^3, t\cos(\pi t)\rangle$ as t ranges from 0 to 1. Find the work done by the force on the object.

Solution. The straightforward way to do this involves substituting the components of $\mathbf{r}$ into $\mathbf{F}$, forming the dot product $\mathbf{F} \cdot \mathbf{r}'$, and then trying to compute the integral, but this integral is extraordinarily messy, perhaps impossible to compute. From Equation 16.1 we know that $\mathbf{F} = \nabla(1/\sqrt{x^2+y^2+z^2})$ so we need only substitute:

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \left.\frac{1}{\sqrt{x^2+y^2+z^2}}\right|_{(1,0,0)}^{(2,1,-1)} = \frac{1}{\sqrt{6}} - 1.$$

♣

Another immediate consequence of the Fundamental Theorem involves **closed paths**. A path C is closed if it forms a loop, so that travelling over the C curve brings you back to the starting point. If C is a closed path, we can integrate around it starting at any point $\mathbf{a}$; since the starting and ending points are the same,

$$\int_C \nabla f \cdot d\mathbf{r} = f(\mathbf{a}) - f(\mathbf{a}) = 0.$$

For example, in a gravitational field (an inverse square law field) the amount of work required to move an object around a closed path is zero. Of course, it's only the *net* amount of work that is zero. It may well take a great deal of work to get from point $\mathbf{a}$ to point $\mathbf{b}$, but then the return trip will "produce" work. For example, it takes work to pump water from a lower to a higher elevation, but if you then let gravity pull the water back down, you can recover work by running a water wheel or generator. (In the real world you won't recover all the work because of various losses along the way.)

To make use of the Fundamental Theorem of Line Integrals, we need to be able to spot conservative vector fields $\mathbf{f}$ and to compute f so that $\mathbf{f} = \nabla f$. Suppose that $\mathbf{f} = \langle f_1, f_2 \rangle = \nabla f$. Then $f_1 = f_x$ and $f_2 = f_y$, and provided that f is sufficiently nice, we know from Clairaut's Theorem (13.23) that $\frac{\partial}{\partial y} f_1 = f_{xy} = f_{yx} = \frac{\partial}{\partial x} f_2$. If we compute $\frac{\partial}{\partial y} f_1$ and $\frac{\partial}{\partial x} f_2$ and find that they are not equal, then $\mathbf{f}$ is not conservative. If $\frac{\partial}{\partial y} f_1 = \frac{\partial}{\partial x} f_2$, then, again provided that $\mathbf{f}$ is sufficiently nice, we can be assured that $\mathbf{f}$ is conservative. Ultimately, what's important is that we be able to find f; as this amounts to finding anti-derivatives, we may not always succeed.

Example 16.11:

Find an f so that $\langle 4x^3 - y^2, 4 - 2xy \rangle = \nabla f$.

Solution. First, note that

$$\frac{\partial}{\partial y}(4x^3 - y^2) = -2y \qquad \text{and} \qquad \frac{\partial}{\partial x}(4 - 2xy) = -2y,$$

so the desired f does exist. This means that $f_x = 4x^3 - y^2$, so that $f = x^4 - xy^2 + g(y)$; the first two terms are needed to get $4x^3 - y^2$, and the $g(y)$ could be any function of y, as it would disappear upon taking a derivative with respect to x. Likewise, since $f_y = 4 - 2xy$, $f = 4y - xy^2 + h(x)$. The question now becomes, is it possible to find $g(y)$ and $h(x)$ so that

$$x^4 - xy^2 + g(y) = 4y - xy^2 + h(x),$$

and of course the answer is yes: $g(y) = 4y$, $h(x) = x^4$. Thus, $f = 4y + x^4 - xy^2$. ♣

We can test a vector field $\mathbf{f} = \langle f_1, f_2, f_3 \rangle$ in a similar way. Suppose that $\langle f_1, f_2, f_3 \rangle = \langle f_x, f_y, f_z \rangle$. If we temporarily hold z constant, then $f(x,y,z)$ is a function of x and y, and by Clairaut's Theorem $\frac{\partial}{\partial y} f_1 =$

$f_{xy} = f_{yx} = \frac{\partial}{\partial x} f_2$. Likewise, holding y constant implies $\frac{\partial}{\partial z} f_1 = f_{xz} = f_{zx} = \frac{\partial}{\partial x} f_3$, and with x constant we get $\frac{\partial}{\partial z} f_2 = f_{yz} = f_{zy} = \frac{\partial}{\partial y} f_3$. Conversely, if we find that $\frac{\partial}{\partial y} f_1 = \frac{\partial}{\partial x} f_2$, $\frac{\partial}{\partial z} f_1 = \frac{\partial}{\partial x} f_3$, and $\frac{\partial}{\partial z} f_2 = \frac{\partial}{\partial y} f_3$ then **f** is conservative.

Exercises for 16.3.3

Exercise 16.3.22 *Find an f so that $\nabla f = \langle 2x + y^2, 2y + x^2 \rangle$, or explain why there is no such f.*

Exercise 16.3.23 *Find an f so that $\nabla f = \langle x^3, -y^4 \rangle$, or explain why there is no such f.*

Exercise 16.3.24 *Find an f so that $\nabla f = \langle xe^y, ye^x \rangle$, or explain why there is no such f.*

Exercise 16.3.25 *Find an f so that $\nabla f = \langle y \cos x, y \sin x \rangle$, or explain why there is no such f.*

Exercise 16.3.26 *Find an f so that $\nabla f = \langle y \cos x, \sin x \rangle$, or explain why there is no such f.*

Exercise 16.3.27 *Find an f so that $\nabla f = \langle x^2 y^3, xy^4 \rangle$, or explain why there is no such f.*

Exercise 16.3.28 *Find an f so that $\nabla f = \langle yz, xz, xy \rangle$, or explain why there is no such f.*

Exercise 16.3.29 *Evaluate $\int_C (10x^4 - 2xy^3)\, dx - 3x^2 y^2\, dy$ where C is the part of the curve $x^5 - 5x^2 y^2 - 7x^2 = 0$ from $(0,0)$ to $(3,2)$.*

Exercise 16.3.30 *Let $\mathbf{F} = \langle yz, xz, xy \rangle$. Find the work done by this force field on an object that moves from $(1,0,2)$ to $(1,2,3)$.*

Exercise 16.3.31 *Let $\mathbf{F} = \langle e^y, xe^y + \sin z, y \cos z \rangle$. Find the work done by this force field on an object that moves from $(0,0,0)$ to $(1,-1,3)$.*

Exercise 16.3.32 *Let*

$$\mathbf{F} = \left\langle \frac{-x}{(x^2+y^2+z^2)^{3/2}}, \frac{-y}{(x^2+y^2+z^2)^{3/2}}, \frac{-z}{(x^2+y^2+z^2)^{3/2}} \right\rangle.$$

Find the work done by this force field on an object that moves from $(1,1,1)$ to $(4,5,6)$.

16.4 Green's Theorem

We now come to the first of three important theorems that extend the Fundamental Theorem of Calculus to higher dimensions. (The Fundamental Theorem of Line Integrals has already done this in one way, but in that case we were still dealing with an essentially one-dimensional integral.) They all share with the Fundamental Theorem the following rather vague description: *To compute a certain sort of integral over a region, we may do a computation on the boundary of the region that involves one fewer integrations.*

Note that this does indeed describe the Fundamental Theorem of Calculus and the Fundamental Theorem of Line Integrals: to compute a single integral over an interval, we do a computation on the boundary (the endpoints) that involves one fewer integrations, namely, no integrations at all.

Theorem 16.12: Green's Theorem

If the vector field $\mathbf{f} = \langle f_1, f_2 \rangle$ and the region D are sufficiently nice, and if C is the boundary of D (C is a closed curve), then

$$\iint_D \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\, dA = \int_C f_1\, dx + f_@\, dy,$$

provided the integration on the right is done counter-clockwise around C.

The proof of Green's Theorem will follow a discussion and several examples.

To indicate that an integral $\int_C$ is being done over a closed curve in the counter-clockwise direction, we usually write $\oint_C$. We also use the notation ∂D to mean the boundary of D **oriented** in the counterclockwise direction. With this notation, $\oint_C = \int_{\partial D}$.

We already know one case, not particularly interesting, in which this theorem is true: If $\mathbf{f}$ is conservative, we know that the integral $\oint_C \mathbf{f} \cdot d\mathbf{r} = 0$, because any integral of a conservative vector field around a closed curve is zero. We also know in this case that $\partial f_1/\partial y = \partial f_2/\partial x$, so the double integral in the theorem is simply the integral of the zero function, namely, 0. So in the case that $\mathbf{f}$ is conservative, the theorem says simply that $0 = 0$.

Example 16.13:

We illustrate the theorem by computing both sides of

$$\int_{\partial D} x^3\, dx + xy^2\, dy = \iint_D y - 0\, dA,$$

where D is the triangular region with corners $(0,0)$, $(1,0)$, $(0,-1)$.

Solution. Starting with the double integral:

$$\iint_D y^2 - 0\, dA = \int_0^1 \int_0^{-1+x} y^2\, dy\, dx = \int_0^1 \frac{(-1+x)^3}{3}\, dx = \left.\frac{(-1+x)^4}{12}\right|_0^1 = -\frac{1}{12}.$$

There is no single formula to describe the boundary of D, so to compute the left side directly we need to compute three separate integrals corresponding to the three sides of the triangle, and each of these integrals we break into two integrals, the "dx" part and the "dy" part. The three sides are described by $y=0$, $y=-1+x$, and $x=0$. The integrals are then

$$\begin{aligned}\int_{\partial D} x^3\,dx+xy^2\,dy &= \int_0^1 x^3\,dx+\int_0^0 0\,dy+\int_1^0 x^3\,dx+\int_0^{-1}(y+1)y^2\,dy+\int_0^0 0\,dx+\int_0^{-1} 0\,dy\\ &= \frac{1}{4}+0-\frac{1}{4}+-\frac{1}{12}+0+0=-\frac{1}{12}.\end{aligned}$$

Alternately, we could describe the three sides in vector form as $\langle t,0\rangle$ for t from 0 to 1, $\langle t+1,t\rangle$ for t from 0 to -1, and $\langle 0,-1+t\rangle$ for t from 0 to 1. Note that in each case, as t ranges from lower to upper bound, we follow the corresponding side in the correct direction. Now

$$\begin{aligned}\int_{\partial D} x^3\,dx+xy^2\,dy &= \int_0^1 t^3+t\cdot 0^2\,dt+\int_0^{-1}(t+1)^3+(t+1)t^2\,dt+\int_0^1 0+0(-1+t)\,dt\\ &= \int_0^1 t^3\,dt+\int_0^{-1}(t+1)^3+(t+1)t^2\,dt=-\frac{1}{12}.\end{aligned}$$

♣

In this case, none of the integrations are difficult, but the second approach is somewhat tedious because of the necessity to set up three different integrals. In different circumstances, either of the integrals, the single or the double, might be easier to compute. Sometimes it is worthwhile to turn a single integral into the corresponding double integral, sometimes exactly the opposite approach is best.

Here is a clever use of Green's Theorem: We know that areas can be computed using double integrals, namely,

$$\iint_D 1\,dA$$

computes the area of region D. If we can find f_1 and f_2 so that $\partial f_2/\partial x-\partial f_1/\partial y=1$, then the area is also

$$\int_{\partial D} f_1\,dx+f_2\,dy.$$

It is quite easy to do this: $f_1=0, f_2=x$ works, as do $f_1=-y, f_2=0$ and $f_1=-y/2, f_2=x/2$.

Example 16.14:

An ellipse centered at the origin, with its two principal axes aligned with the x and y axes, is given by

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

Find the area of the interior of the ellipse.

Solution. We find the area of the interior of the ellipse via Green's theorem. To do this we need a vector equation for the boundary; one such equation is $\langle a\cos t, b\sin t\rangle$, as t ranges from 0 to 2π. We can easily verify this by substitution:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{a^2\cos^2 t}{a^2}+\frac{b^2\sin^2 t}{b^2}=\cos^2 t+\sin^2 t=1.$$

Let's consider the three possibilities for f_1 and f_2 above: Using 0 and x gives

$$\oint_C 0\,dx + x\,dy = \int_0^{2\pi} a\cos(t) b\cos(t)\,dt = \int_0^{2\pi} ab\cos^2(t)\,dt.$$

Using $-y$ and 0 gives

$$\oint_C -y\,dx + 0\,dy = \int_0^{2\pi} -b\sin(t)(-a\sin(t))\,dt = \int_0^{2\pi} ab\sin^2(t)\,dt.$$

Finally, using $-y/2$ and $x/2$ gives

$$\begin{aligned}\oint_C -\frac{y}{2}\,dx + \frac{x}{2}\,dy &= \int_0^{2\pi} -\frac{b\sin(t)}{2}(-a\sin(t))\,dt + \frac{a\cos(t)}{2}(b\cos(t))\,dt \\ &= \int_0^{2\pi} \frac{ab\sin^2 t}{2} + \frac{ab\cos^2 t}{2}\,dt = \int_0^{2\pi} \frac{ab}{2}\,dt = \pi ab.\end{aligned}$$

The first two integrals are not particularly difficult, but the third is very easy, though the choice of f_1 and f_2 seems more complicated. ♣

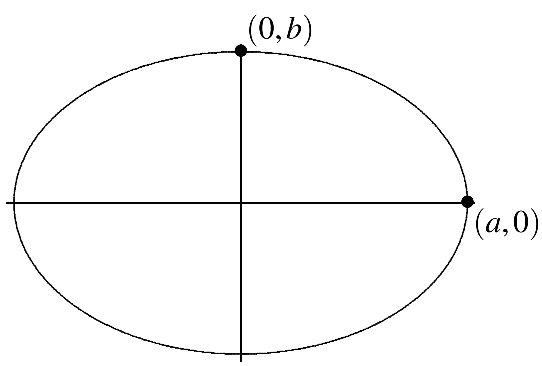


Figure 16.3: A "standard" ellipse, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Now we look at the proof of Green's Theorem.

Proof. We cannot here prove Green's Theorem in general, but we can do a special case. We seek to prove that for a vector field $\mathbf{f} = \langle f_1, f_2 \rangle$

$$\oint_C f_1\,dx + f_2\,dy = \iint_D \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\,dA.$$

It is sufficient to show that

$$\oint_C f_1\,dx = \iint_D -\frac{\partial f_1}{\partial y}\,dA \qquad \text{and} \qquad \oint_C f_2\,dy = \iint_D \frac{\partial f_2}{\partial x}\,dA,$$

which we can do if we can compute the double integral in both possible ways, that is, using $dA = dy\,dx$ and $dA = dx\,dy$.

For the first equation, we start with

$$\iint_D \frac{\partial f_1}{\partial y}\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} \frac{\partial f_1}{\partial y}\,dy\,dx = \int_a^b f_1(x, g_2(x)) - f_1(x, g_1(x))\,dx.$$

Here we have simply used the ordinary Fundamental Theorem of Calculus, since for the inner integral we are integrating a derivative with respect to y: an antiderivative of $\partial f_1/\partial y$ with respect to y is simply $f_1(x,y)$, and then we substitute g_1 and g_2 for y and subtract.

Now we need to manipulate $\oint_C f_1\,dx$. The boundary of region D consists of 4 parts, given by the equations $y = g_1(x)$, $x = b$, $y = g_2(x)$, and $x = a$. On the portions $x = b$ and $x = a$, $dx = 0\,dt$, so the corresponding integrals are zero. For the other two portions, we use the parametric forms $x = t$, $y = g_1(t)$, $a \le t \le b$, and $x = t$, $y = g_2(t)$, letting t range from b to a, since we are integrating counter-clockwise around the boundary. The resulting integrals give us

$$\begin{aligned}\oint_C f_1\,dx = \int_a^b f_1(t,g_1(t))\,dt + \int_b^a f_1(t,g_2(t))\,dt &= \int_a^b f_1(t,g_1(t))\,dt - \int_a^b f_1(t,g_2(t))\,dt\\ &= \int_a^b f_1(t,g_1(t)) - f_1(t,g_2(t))\,dt\end{aligned}$$

which is the result of the double integral times -1, as desired.

The equation involving f_2 is essentially the same, and left as an exercise. ♣

We can now rewrite Green's Theorem using the concepts of divergence and curl; these rewritten versions in turn are closer to some later theorems we will see.

Suppose we write a two dimensional vector field in the form $\mathbf{f} = \langle f_1, f_2, 0\rangle$, where f_1 and f_2 are functions of x and y. Then

$$\nabla\times\mathbf{f} = \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ f_1 & f_2 & 0\end{vmatrix} = \langle 0,0,\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\rangle,$$

and so $(\nabla\times\mathbf{f})\cdot\mathbf{k} = \langle 0,0,\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\rangle\cdot\langle 0,0,1\rangle = \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}$. So Green's Theorem says

$$\int_{\partial D}\mathbf{f}\cdot d\mathbf{r} = \int_{\partial D} f_1\,dx + f_2 dy = \iint_D \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\,dA = \iint_D (\nabla\times\mathbf{f})\cdot\mathbf{k}\,dA. \tag{16.2}$$

Roughly speaking, the right-most integral adds up the curl (tendency to swirl) at each point in the region; the left-most integral adds up the tangential components of the vector field around the entire boundary. Green's Theorem says these are equal, or roughly, that the sum of the "microscopic" swirls over the region is the same as the "macroscopic" swirl around the boundary.

Next, suppose that the boundary ∂D has a vector form $\mathbf{r}(t)$, so that $\mathbf{r}'(t)$ is tangent to the boundary, and $\mathbf{T} = \mathbf{r}'(t)/|\mathbf{r}'(t)|$ is the usual unit tangent vector. Writing $\mathbf{r} = \langle x(t), y(t)\rangle$ we get

$$\mathbf{T} = \frac{\langle x', y'\rangle}{|\mathbf{r}'(t)|}$$

and then

$$\mathbf{N} = \frac{\langle y', -x'\rangle}{|\mathbf{r}'(t)|}$$

is a unit vector perpendicular to $\mathbf{T}$, that is, a unit normal to the boundary. Now

$$\int_{\partial D}\mathbf{f}\cdot\mathbf{N}\,ds = \int_{\partial D}\langle f_1, f_2\rangle\cdot\frac{\langle y', -x'\rangle}{|\mathbf{r}'(t)|}|\mathbf{r}'(t)|dt = \int_{\partial D} f_1 y'\,dt - f_2 x'\,dt$$

$$= \int_{\partial D} f_1\,dy - f_2\,dx = \int_{\partial D} -f_2\,dx + f_1\,dy.$$

So far, we've just rewritten the original integral using alternate notation. The last integral looks just like the left side of Green's Theorem (16.12) except that f_1 and f_2 have traded places and f_2 has acquired a negative sign. Then applying Green's Theorem we get

$$\int_{\partial D} -f_2\,dx + f_1\,dy = \iint_D \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y}\,dA = \iint_D \nabla \cdot \mathbf{f}\,dA.$$

Summarizing the long string of equalities,

$$\int_{\partial D} \mathbf{f} \cdot \mathbf{N}\,ds = \iint_D \nabla \cdot \mathbf{f}\,dA. \tag{16.3}$$

Roughly speaking, the first integral adds up the flow across the boundary of the region, from inside to out, and the second sums the divergence (tendency to spread) at each point in the interior. The theorem roughly says that the sum of the "microscopic" spreads is the same as the total spread across the boundary and out of the region.

Exercises for 16.4

Exercise 16.4.1 *Compute* $\int_{\partial D} 2y\,dx + 3x\,dy$, *where* D *is described by* $0 \le x \le 1$, $0 \le y \le 1$.

Exercise 16.4.2 *Compute* $\int_{\partial D} xy\,dx + xy\,dy$, *where* D *is described by* $0 \le x \le 1$, $0 \le y \le 1$.

Exercise 16.4.3 *Compute* $\int_{\partial D} e^{2x+3y}\,dx + e^{xy}\,dy$, *where* D *is described by* $-2 \le x \le 2$, $-1 \le y \le 1$.

Exercise 16.4.4 *Compute* $\int_{\partial D} y\cos x\,dx + y\sin x\,dy$, *where* D *is described by* $0 \le x \le \pi/2$, $1 \le y \le 2$.

Exercise 16.4.5 *Compute* $\int_{\partial D} x^2 y\,dx + xy^2\,dy$, *where* D *is described by* $0 \le x \le 1$, $0 \le y \le x$.

Exercise 16.4.6 *Compute* $\int_{\partial D} x\sqrt{y}\,dx + \sqrt{x+y}\,dy$, *where* D *is described by* $1 \le x \le 2$, $2x \le y \le 4$.

Exercise 16.4.7 *Compute* $\int_{\partial D} (x/y)\,dx + (2+3x)\,dy$, *where* D *is described by* $1 \le x \le 2$, $1 \le y \le x^2$.

Exercise 16.4.8 *Compute* $\int_{\partial D} \sin y\,dx + \sin x\,dy$, *where* D *is described by* $0 \le x \le \pi/2$, $x \le y \le \pi/2$.

Exercise 16.4.9 *Compute* $\int_{\partial D} x\ln y\,dx$, *where* D *is described by* $1 \le x \le 2$, $e^x \le y \le e^{x^2}$.

Exercise 16.4.10 *Compute* $\int_{\partial D} \sqrt{1+x^2}\, dy$*, where* D *is described by* $-1 \le x \le 1$*,* $x^2 \le y \le 1$*.*

Exercise 16.4.11 *Compute* $\int_{\partial D} x^2 y\, dx - xy^2\, dy$*, where* D *is described by* $x^2 + y^2 \le 1$*.*

Exercise 16.4.12 *Compute* $\int_{\partial D} y^3\, dx + 2x^3\, dy$*, where* D *is described by* $x^2 + y^2 \le 4$*.*

Exercise 16.4.13 *Evaluate* $\oint_C (y - \sin(x))\, dx + \cos(x)\, dy$*, where* C *is the boundary of the triangle with vertices* $(0,0)$*,* $(1,0)$*, and* $(1,2)$ *oriented counter-clockwise.*

Exercise 16.4.14 *Finish our proof of Green's Theorem by showing that* $\oint_C f_2\, dy = \iint_D \frac{\partial f_2}{\partial x}\, dA$*.*

16.5 The Divergence Theorem

The third version of Green's Theorem (Equation 16.3) we saw was:

$$\int_{\partial D} \mathbf{f} \cdot \mathbf{N}\, ds = \iint_D \nabla \cdot \mathbf{f}\, dA.$$

With minor changes this turns into another equation, the Divergence Theorem:

Theorem 16.15: Divergence Theorem

Under suitable conditions, if E *is a region of three dimensional space and* D *is its boundary surface, oriented outward, then*

$$\iint_D \mathbf{f} \cdot \mathbf{N}\, dS = \iiint_E \nabla \cdot \mathbf{f}\, dV.$$

Proof. Again this theorem is too difficult to prove here, but a special case is easier. In the proof of a special case of Green's Theorem, we needed to know that we could describe the region of integration in both possible orders, so that we could set up one double integral using $dx\,dy$ and another using $dy\,dx$. Similarly here, we need to be able to describe the three-dimensional region E in different ways.

We start by rewriting the triple integral:

$$\iiint_E \nabla \cdot \mathbf{f}\, dV = \iiint_E \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}\right) dV = \iiint_E \frac{\partial f_1}{\partial x}\, dV + \iiint_E \frac{\partial f_2}{\partial y}\, dV + \iiint_E \frac{\partial f_3}{\partial z}\, dV.$$

The double integral may be rewritten:

$$\iint_D \mathbf{f} \cdot \mathbf{N}\, dS = \iint_D (f_1\mathbf{i} + f_2\mathbf{j} + f_3\mathbf{k}) \cdot \mathbf{N}\, dS = \iint_D f_1\mathbf{i} \cdot \mathbf{N}\, dS + \iint_D f_2\mathbf{j} \cdot \mathbf{N}\, dS + \iint_D f_3\mathbf{k} \cdot \mathbf{N}\, dS.$$

To prove that these give the same value it is sufficient to prove that

$$\iint_D f_1\mathbf{i}\cdot\mathbf{N}\,dS = \iiint_E \frac{\partial f_1}{\partial x}\,dV, \tag{16.4}$$

$$\iint_D f_2\mathbf{j}\cdot\mathbf{N}\,dS = \iiint_E \frac{\partial f_2}{\partial y}\,dV, \text{ and} \tag{16.5}$$

$$\iint_D f_3\mathbf{k}\cdot\mathbf{N}\,dS = \iiint_E \frac{\partial f_3}{\partial z}\,dV. \tag{16.6}$$

Not surprisingly, these are all pretty much the same; we'll do the first one.

We set the triple integral up with dx innermost:

$$\iiint_E \frac{\partial f_1}{\partial x}\,dV = \iint_B \int_{g_1(y,z)}^{g_2(y,z)} \frac{\partial f_1}{\partial x}\,dx\,dA = \iint_B f_1(g_2(y,z),y,z) - f_1(g_1(y,z),y,z)\,dA,$$

where B is the region in the y-z plane over which we integrate. The boundary surface of E consists of a "top" $x = g_2(y,z)$, a "bottom" $x = g_1(y,z)$, and a "wrap-around side" that is vertical to the y-z plane. To integrate over the entire boundary surface, we can integrate over each of these (top, bottom, side) and add the results. Over the side surface, the vector $\mathbf{N}$ is perpendicular to the vector $\mathbf{i}$, so

$$\iint_{\text{side}} f_1\mathbf{i}\cdot\mathbf{N}\,dS = \iint_{\text{side}} 0\,dS = 0.$$

Thus, we are left with just the surface integral over the top plus the surface integral over the bottom. For the top, we use the vector function $\mathbf{r} = \langle g_2(y,z),y,z\rangle$ which gives $\mathbf{r}_y\times\mathbf{r}_z = \langle 1,-g_{2y},-g_{2z}\rangle$; the dot product of this with $\mathbf{i} = \langle 1,0,0\rangle$ is 1. Then

$$\iint_{\text{top}} f_1\mathbf{i}\cdot\mathbf{N}\,dS = \iint_B f_1(g_2(y,z),y,z)\,dA.$$

In almost identical fashion we get

$$\iint_{\text{bottom}} f_1\mathbf{i}\cdot\mathbf{N}\,dS = -\iint_B f_1(g_1(y,z),y,z)\,dA,$$

where the negative sign is needed to make $\mathbf{N}$ point in the negative x direction. Now

$$\iint_D f_1\mathbf{i}\cdot\mathbf{N}\,dS = \iint_B f_1(g_2(y,z),y,z)\,dA - \iint_B f_1(g_1(y,z),y,z)\,dA,$$

which is the same as the value of the triple integral above. ♣

It is worth noting that this theorem is also referred to as *Gauss' Theorem*. We now compute an example.

Example 16.16:

Let $\mathbf{f} = \langle 2x,3y,z^2\rangle$, and consider the three-dimensional volume inside the cube with faces parallel to the principal planes and opposite corners at $(0,0,0)$ and $(1,1,1)$. Compute the two integrals of the divergence theorem.

Solution. The triple integral is the easier of the two:

$$\int_0^1\int_0^1\int_0^1 2+3+2z\,dx\,dy\,dz = 6.$$

The surface integral must be separated into six parts, one for each face of the cube. One face is $z=0$ or $\mathbf{r}=\langle u,v,0\rangle$, $0\le u,v\le 1$. Then $\mathbf{r}_u=\langle 1,0,0\rangle$, $\mathbf{r}_v=\langle 0,1,0\rangle$, and $\mathbf{r}_u\times\mathbf{r}_v=\langle 0,0,1\rangle$. We need this to be oriented downward (out of the cube), so we use $\langle 0,0,-1\rangle$ and the corresponding integral is

$$\int_0^1\int_0^1 -z^2\,du\,dv = \int_0^1\int_0^1 0\,du\,dv = 0.$$

Another face is $y=1$ or $\mathbf{r}=\langle u,1,v\rangle$. Then $\mathbf{r}_u=\langle 1,0,0\rangle$, $\mathbf{r}_v=\langle 0,0,1\rangle$, and $\mathbf{r}_u\times\mathbf{r}_v=\langle 0,-1,0\rangle$. We need a normal in the positive y direction, so we convert this to $\langle 0,1,0\rangle$, and the corresponding integral is

$$\int_0^1\int_0^1 3y\,du\,dv = \int_0^1\int_0^1 3\,du\,dv = 3.$$

The remaining four integrals have values 0, 0, 2, and 1, and the sum of these is 6, in agreement with the triple integral. ♣

Example 16.17:

Let $\mathbf{f}=\langle x^3,y^3,z^2\rangle$, and consider the cylindrical volume $x^2+y^2\le 9$, $0\le z\le 2$. Compute the two integrals of the divergence theorem.

Solution. The triple integral (using cylindrical coordinates) is

$$\int_0^{2\pi}\int_0^3\int_0^2 (3r^2+2z)r\,dz\,dr\,d\theta = 279\pi.$$

For the surface we need three integrals. The top of the cylinder can be represented by $\mathbf{r}=\langle v\cos u, v\sin u, 2\rangle$; $\mathbf{r}_u\times\mathbf{r}_v=\langle 0,0,-v\rangle$, which points down into the cylinder, so we convert it to $\langle 0,0,v\rangle$. Then

$$\int_0^{2\pi}\int_0^3 \langle v^3\cos^3 u, v^3\sin^3 u, 4\rangle\cdot\langle 0,0,v\rangle\,dv\,du = \int_0^{2\pi}\int_0^3 4v\,dv\,du = 36\pi.$$

The bottom is $\mathbf{r}=\langle v\cos u, v\sin u, 0\rangle$; $\mathbf{r}_u\times\mathbf{r}_v=\langle 0,0,-v\rangle$ and

$$\int_0^{2\pi}\int_0^3 \langle v^3\cos^3 u, v^3\sin^3 u, 0\rangle\cdot\langle 0,0,-v\rangle\,dv\,du = \int_0^{2\pi}\int_0^3 0\,dv\,du = 0.$$

The side of the cylinder is $\mathbf{r}=\langle 3\cos u, 3\sin u, v\rangle$; $\mathbf{r}_u\times\mathbf{r}_v=\langle 3\cos u, 3\sin u, 0\rangle$ which does point outward, so

$$\begin{aligned}\int_0^{2\pi}\int_0^2 &\langle 27\cos^3 u, 27\sin^3 u, v^2\rangle\cdot\langle 3\cos u, 3\sin u, 0\rangle\,dv\,du\\ &= \int_0^{2\pi}\int_0^2 81\cos^4 u + 81\sin^4 u\,dv\,du = 243\pi.\end{aligned}$$

The total surface integral is thus $36\pi+0+243\pi=279\pi$. ♣

Exercises for 16.5

Exercise 16.5.1 *Using* $\boldsymbol{f} = \langle 3x, y^3, -2z^2 \rangle$ *and the region bounded by* $x^2 + y^2 = 9$, $z = 0$, *and* $z = 5$, *compute both integrals from the Divergence Theorem.*

Exercise 16.5.2 *Let E be the volume described by* $0 \le x \le a$, $0 \le y \le b$, $0 \le z \le c$, *and* $\boldsymbol{f} = \langle x^2, y^2, z^2 \rangle$. *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.3 *Let E be the volume described by* $0 \le x \le 1$, $0 \le y \le 1$, $0 \le z \le 1$, *and* $\boldsymbol{f} = \langle 2xy, 3xy, ze^{x+y} \rangle$. *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.4 *Let E be the volume described by* $0 \le x \le 1$, $0 \le y \le x$, $0 \le z \le x + y$, *and* $\boldsymbol{f} = \langle x, 2y, 3z \rangle$. *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.5 *Let E be the volume described by* $x^2 + y^2 + z^2 \le 4$, *and* $\boldsymbol{f} = \langle x^3, y^3, z^3 \rangle$. *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.6 *Let E be the hemisphere described by* $0 \le z \le \sqrt{1 - x^2 - y^2}$, *and* $\boldsymbol{f} = \langle \sqrt{x^2 + y^2 + z^2}, \sqrt{x^2 + y^2 + z^2}, \sqrt{}$ *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.7 *Let E be the volume described by* $x^2 + y^2 \le 1$, $0 \le z \le 4$, *and* $\boldsymbol{f} = \langle xy^2, yz, x^2 z \rangle$. *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.8 *Let E be the solid cone above the x-y plane and inside* $z = 1 - \sqrt{x^2 + y^2}$, *and* $\boldsymbol{f} = \langle x\cos^2 z, y\sin^2 z, \sqrt{x^2 + y^2} z \rangle$. *Compute* $\iint_{\partial E} \boldsymbol{f} \cdot \boldsymbol{N}\, dS$.

Exercise 16.5.9 *Prove the other two equations in the display 16.4.*

Exercise 16.5.10 *Suppose D is a closed surface, and that D and F are sufficiently nice. Show that*

$$\iint_D (\nabla \times \boldsymbol{f}) \cdot \boldsymbol{N}\, dS = 0$$

where $\boldsymbol{N}$ *is the outward pointing unit normal.*

Exercise 16.5.11 *Suppose D is a closed surface, D is sufficiently nice, and* $F = \langle a, b, c \rangle$ *is a constant vector field. Show that*

$$\iint_D \boldsymbol{f} \cdot \boldsymbol{N}\, dS = 0$$

where $\boldsymbol{N}$ *is the outward pointing unit normal.*

Exercise 16.5.12 *We know that the volume of a region E may often be computed as $\iiint_E dx\,dy\,dz$. Show that this volume may also be computed as $\frac{1}{3}\iint_{\partial E}\langle x,y,z\rangle\cdot\mathbf{N}\,dS$ where $\mathbf{N}$ is the outward pointing unit normal to ∂E.*

16.6 Vector Functions for Surfaces

We have dealt extensively with vector equations for curves, $\mathbf{r}(t)=\langle x(t),y(t),z(t)\rangle$. A similar technique can be used to represent surfaces in a way that is more general than the equations for surfaces we have used so far. Recall that when we use $\mathbf{r}(t)$ to represent a curve, we imagine the vector $\mathbf{r}(t)$ with its tail at the origin, and then we follow the head of the arrow as t changes. The vector "draws" the curve through space as t varies.

Suppose we instead have a vector function of two variables,

$$\mathbf{r}(u,v)=\langle x(u,v),y(u,v),z(u,v)\rangle.$$

As both u and v vary, we again imagine the vector $\mathbf{r}(u,v)$ with its tail at the origin, and its head sweeps out a surface in space. A useful analogy is the technology of CRT video screens, in which an electron gun fires electrons in the direction of the screen. The gun's direction sweeps horizontally and vertically to "paint" the screen with the desired image. In practice, the gun moves horizontally through an entire line, then moves vertically to the next line and repeats the operation. In the same way, it can be useful to imagine fixing a value of v and letting $\mathbf{r}(u,v)$ sweep out a curve as u changes. Then v can change a bit, and $\mathbf{r}(u,v)$ sweeps out a new curve very close to the first. Put enough of these curves together and they form a surface.

Example 16.18:

Consider the function $\mathbf{r}(u,v)=\langle v\cos u, v\sin u, v\rangle$. For a fixed value of v, as u varies from 0 to 2π, this traces a circle of radius v at height v above the x-y plane. Put lots and lots of these together,and they form a cone, as in Figure 16.4.

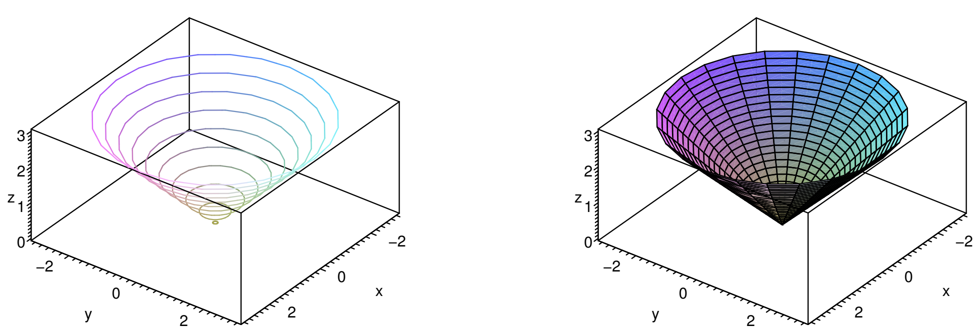


Figure 16.4: Tracing a surface.

Recall that from the vector equation of the curve we can compute the unit tangent $\mathbf{T}$, the unit normal $\mathbf{N}$, and the binormal vector $\mathbf{B} = \mathbf{T} \times \mathbf{N}$; you may want to review Section 15.3. The binormal is perpendicular to both $\mathbf{T}$ and $\mathbf{N}$; one way to interpret this is that $\mathbf{N}$ and $\mathbf{B}$ define a plane perpendicular to $\mathbf{T}$, that is, perpendicular to the curve; since $\mathbf{N}$ and $\mathbf{B}$ are perpendicular to each other, they can function just as $\mathbf{i}$ and $\mathbf{j}$ do for the x-y plane.

Example 16.19:

The curve given by

$$\mathbf{r} = \langle (2+\cos(3u/2))\cos u, (2+\cos(3u/2))\sin u, \sin(3u/2)\rangle$$

is called a trefoil knot. Then $\mathbf{c}(v) = \mathbf{N}\cos v + \mathbf{B}\sin v$ is a vector equation for a unit circle in a plane perpendicular to the curve described by $\mathbf{r}$, except that the usual interpretation of $\mathbf{c}$ would put its center at the origin. We can fix that simply by adding $\mathbf{c}$ to the original $\mathbf{r}$: let $\mathbf{f} = \mathbf{r}(u) + \mathbf{c}(v)$. For a fixed u this draws a circle around the point $\mathbf{r}(u)$; as u varies we get a sequence of such circles around the curve $\mathbf{r}$, that is, a tube of radius 1 with $\mathbf{r}$ at its center. We can easily change the radius; for example $\mathbf{r}(u) + a\mathbf{c}(v)$ gives the tube radius a; we can make the radius vary as we move along the curve with $\mathbf{r}(u) + g(u)\mathbf{c}(v)$, where $g(u)$ is a function of u.

As shown in Figure 16.5, it is hard to see that the plain knot is knotted; the tube makes the structure apparent. Of course, there is nothing special about the trefoil knot in this example; we can put a tube around (almost) any curve in the same way.

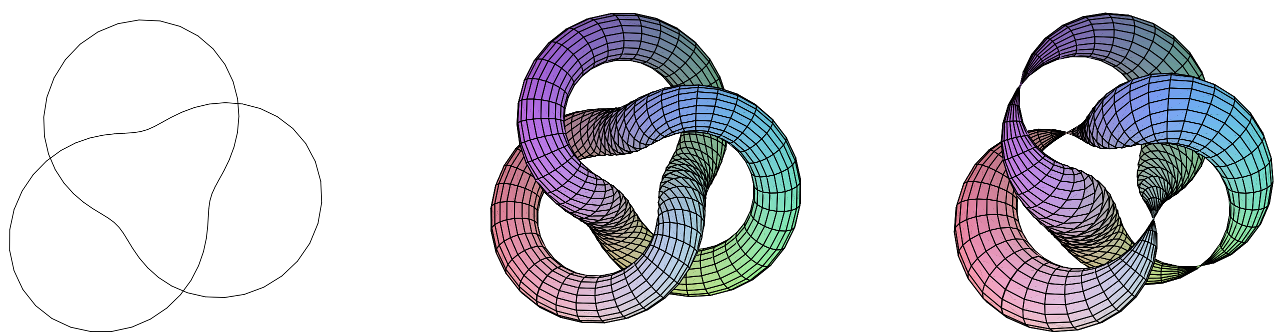


Figure 16.5: Tubes around a trefoil knot, with radius $1/2$ and $3\cos(u)/4$.

We have previously examined surfaces given in the form $f(x,y)$. It is sometimes useful to represent such surfaces in the more general vector form, which is quite easy: $\mathbf{r}(u,v) = \langle u, v, f(u,v)\rangle$. The names of the variables are not important of course; instead of disguising x and y, we could simply write $\mathbf{r}(x,y) = \langle x, y, f(x,y)\rangle$.

We have also previously dealt with surfaces that are not functions of x and y; many of these are easy to represent in vector form. One common type of surface that cannot be represented as $z = f(x,y)$ is a surface given by an equation involving only x and y. For example, $x + y = 1$ and $y = x^2$ are "vertical" surfaces. For every point (x,y) in the plane that satisfies the equation, the point (x,y,z) is on the surface, for every value of z. Thus, a corresponding vector form for the surface is something like $\langle f(u), g(u), v\rangle$; for example, $x + y = 1$ becomes $\langle u, 1-u, v\rangle$ and $y = x^2$ becomes $\langle u, u^2, v\rangle$.

Yet another sort of example is the sphere, say $x^2+y^2+z^2=1$. This cannot be written in the form $z=f(x,y)$, but it is easy to write in vector form; indeed this particular surface is much like the cone, since it has circular cross-sections, or we can think of it as a tube around a portion of the z-axis, with a radius that varies depending on where along the axis we are. One vector expression for the sphere is $\langle\sqrt{1-v^2}\cos u, \sqrt{1-v^2}\sin u, v\rangle$—this emphasizes the tube structure, as it is naturally viewed as drawing a circle of radius $\sqrt{1-v^2}$ around the z-axis at height v. We could also take a cue from spherical coordinates, and write $\langle\sin u\cos v, \sin u\sin v, \cos u\rangle$, where in effect u and v are ϕ and θ in disguise.

It is quite simple to use a computer program to plot any surface for which you have a vector representation. Using different vector functions sometimes gives different looking plots, because the computer in effect draws the surface by holding one variable constant and then the other.

Here's a simple but striking example: the plane $x+y+z=1$ can be represented quite naturally as $\langle u, v, 1-u-v\rangle$. However we could also think of painting the same plane by choosing a particular point on the plane, say $(1,0,0)$, and then drawing circles or ellipses (or any of a number of other curves) as if that point were the origin in the plane. For example, $\langle 1-v\cos u-v\sin u, v\sin u, v\cos u\rangle$ is one such vector function. Note that while it may not be obvious where this came from, it is quite easy to see that the sum of the x, y, and z components of the vector is always 1. Computer renderings of the plane using these two functions are shown in Figure 16.6.

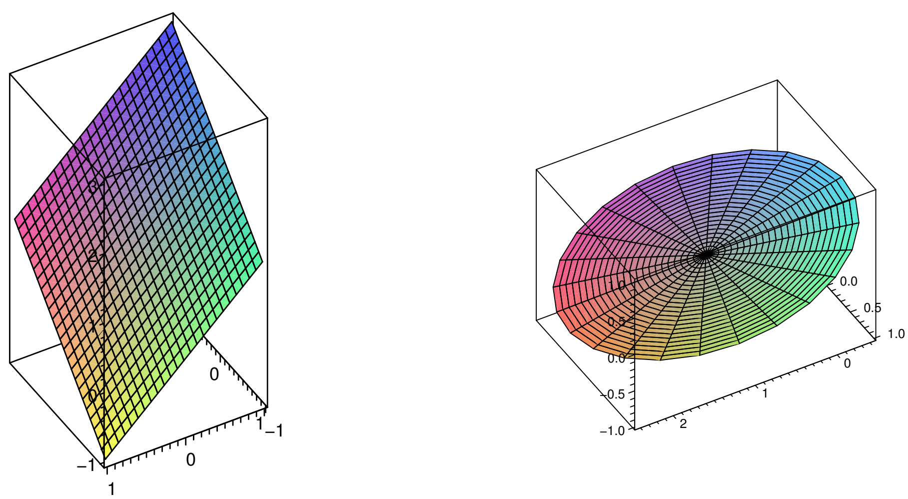


Figure 16.6: Two representations of the same plane.

Suppose we know that a plane contains a particular point (x_0,y_0,z_0) and that two vectors $\mathbf{u}=\langle u_0,u_1,u_2\rangle$ and $\mathbf{v}=\langle v_0,v_1,v_2\rangle$ are parallel to the plane but not to each other. We know how to get an equation for the plane in the form $ax+by+cz=d$, by first computing $\mathbf{u}\times\mathbf{v}$. It's even easier to get a vector equation:

$$\mathbf{r}(u,v)=\langle x_0,y_0,z_0\rangle+u\mathbf{u}+v\mathbf{v}.$$

The first vector gets to the point (x_0,y_0,z_0) and then by varying u and v, $u\mathbf{u}+v\mathbf{v}$ gets to every point in the plane.

Returning to $x+y+z=1$, the points $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ are all on the plane. By subtracting coordinates we see that $\langle -1,0,1\rangle$ and $\langle -1,1,0\rangle$ are parallel to the plane, so a third vector form for this plane is

$$\langle 1,0,0\rangle+u\langle -1,0,1\rangle+v\langle -1,1,0\rangle=\langle 1-u-v, v, u\rangle.$$

This is clearly quite similar to the first form we found.

We have already seen (Section 14.4) how to find the area of a surface when it is defined in the form $f(x,y)$. Finding the area when the surface is given as a vector function is very similar. Looking at the plots of surfaces we have just seen, it is evident that the two sets of curves that fill out the surface divide it into a grid, and that the spaces in the grid are approximately parallelograms. As before this is the key: we can write down the area of a typical small parallelogram and add them all up with an integral.

Suppose we want to approximate the area of the surface $\mathbf{r}(u,v)$ near $\mathbf{r}(u_0,v_0)$. The functions $\mathbf{r}(u,v_0)$ and $\mathbf{r}(u_0,v)$ define two curves that intersect at $\mathbf{r}(u_0,v_0)$. The derivatives of $\mathbf{r}$ give us vectors tangent to these two curves: $\mathbf{r}_u(u_0,v_0)$ and $\mathbf{r}_v(u_0,v_0)$, and then $\mathbf{r}_u(u_0,v_0)\,du$ and $\mathbf{r}_v(u_0,v_0)\,dv$ are two small tangent vectors, whose lengths can be used as the lengths of the sides of an approximating parallelogram. Finally, the area of this parallelogram is $|\mathbf{r}_u \times \mathbf{r}_v|\,du\,dv$ and so the total surface area is

$$\int_a^b \int_c^d |\mathbf{r}_u \times \mathbf{r}_v|\,du\,dv \tag{16.7}$$

Example 16.20:

Find the area of the surface $\langle v\cos u, v\sin u, v\rangle$ *for* $0 \le u \le 2\pi$ *and* $0 \le v \le \frac{1}{2}$.

Solution. We compute $\mathbf{r}_u = \langle -v\sin u, v\cos u, 0\rangle$ and $\mathbf{r}_v = \langle \cos u, \sin u, 1\rangle$. The cross product of these two vectors is $\langle v\cos u, v\sin u, -v\rangle$ with length $\sqrt{2v^2} = \sqrt{2}v$, and the surface area is

$$\int_0^{\pi} \int_0^{1} \sqrt{2}v\,dv\,du = \frac{\sqrt{2}\pi}{4}.$$

♣

Exercises for 16.6

Exercise 16.6.1 *Describe or sketch the surface with the given vector function.*

(a) $\mathbf{r}(u,v) = \langle 6v - u, 3 + v, 1 - 4v\rangle$

(b) $\mathbf{r}(u,v) = \langle 2\cos u, 5\cos u, v\rangle$

(c) $\mathbf{r}(s,t) = \langle u + v, 2u + v, u^2 + v^2\rangle$

(d) $\mathbf{r}(s,t) = \langle \sin u + \cos u, u, v\rangle$

Exercise 16.6.2 *Find the area of the portion of* $x + 2y + 4z = 10$ *in the first octant.*

Exercise 16.6.3 *Find the area of the portion of* $2x + 4y + z = 0$ *inside* $x^2 + y^2 = 1$.

Exercise 16.6.4 *Find the area of $z = x^2 + y^2$ that lies below $z = 1$.*

Exercise 16.6.5 *Find the area of $z = \sqrt{x^2 + y^2}$ that lies below $z = 2$.*

Exercise 16.6.6 *Find the area of the portion of $x^2 + y^2 + z^2 = a^2$ that lies in the first octant.*

Exercise 16.6.7 *Find the area of the portion of $x^2 + y^2 + z^2 = a^2$ that lies above $x^2 + y^2 \le b^2$.*

Exercise 16.6.8 *Find the area of $z = x^2 - y^2$ that lies inside $x^2 + y^2 = a^2$.*

Exercise 16.6.9 *Find the area of $x^2 + y^2 + z^2 = a^2$ that lies above the interior of the circle given in polar coordinates by $r = a\cos\theta$.*

Exercise 16.6.10 *Find the area of the cone $z = k\sqrt{x^2 + y^2}$ that lies above the interior of the circle given in polar coordinates by $r = a\cos\theta$.*

Exercise 16.6.11 *Find the area of the plane $z = ax + by + c$ that lies over a region D with area A.*

Exercise 16.6.12 *Find the area of the cone $z = k\sqrt{x^2 + y^2}$ that lies over a region D with area A.*

Exercise 16.6.13 *Find the area of the cylinder $x^2 + z^2 = a^2$ that lies inside the cylinder $x^2 + y^2 = a^2$.*

Exercise 16.6.14 *The surface $f(x,y)$ can be represented with the vector function $\langle x, y, f(x,y)\rangle$. Set up the surface area integral using this vector function and compare to the integral of Section 14.4.*

16.7 Surface Integrals

In the integral for surface area, from Equation 16.7,

$$\int_a^b \int_c^d |\mathbf{r}_u \times \mathbf{r}_v|\, du\, dv,$$

the integrand $|\mathbf{r}_u \times \mathbf{r}_v|\, du\, dv$ is the area of a tiny parallelogram, that is, a very small surface area, so it is reasonable to abbreviate it dS; then a shortened version of the integral is

$$\iint_D 1 \cdot dS.$$

We have already seen that if D is a region in the plane, the area of D may be computed with

$$\iint_D 1 \cdot dA,$$

so this is really quite familiar, but the dS hides a little more detail than does dA.

Just as we can integrate functions $f(x,y)$ over regions in the plane, using

$$\iint_D f(x,y)\,dA,$$

so we can compute integrals over surfaces in space, using

$$\iint_D f(x,y,z)\,dS.$$

In practice this means that we have a vector function $\mathbf{r}(u,v) = \langle x(u,v), y(u,v), z(u,v)\rangle$ for the surface, and the integral we compute is

$$\int_a^b \int_c^d f(x(u,v), y(u,v), z(u,v))|\mathbf{r}_u \times \mathbf{r}_v|\,du\,dv.$$

That is, we express everything in terms of u and v, and then we can do an ordinary double integral.

Example 16.21: Mass and Center of Mass

Suppose a thin object occupies the upper hemisphere of $x^2+y^2+z^2 = 1$ and has density $\sigma(x,y,z) = z$. Find the mass and center of mass of the object. (Note that the object is just a thin shell; it does not occupy the interior of the hemisphere.)

Solution. We write the hemisphere as $\mathbf{r}(\phi,\theta) = \langle \cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi\rangle$, $0 \le \phi \le \pi/2$ and $0 \le \theta \le 2\pi$. So $\mathbf{r}_\theta = \langle -\sin\theta\sin\phi, \cos\theta\sin\phi, 0\rangle$ and $\mathbf{r}_\phi = \langle \cos\theta\cos\phi, \sin\theta\cos\phi, -\sin\phi\rangle$. Then

$$\mathbf{r}_\theta \times \mathbf{r}_\phi = \langle -\cos\theta\sin^2\phi, -\sin\theta\sin^2\phi, -\cos\phi\sin\phi\rangle$$

and

$$|\mathbf{r}_\theta \times \mathbf{r}_\phi| = |\sin\phi| = \sin\phi,$$

since we are interested only in $0 \le \phi \le \pi/2$. Finally, the density is $z = \cos\phi$ and the integral for mass is

$$\int_0^{2\pi}\int_0^{\pi/2} \cos\phi\sin\phi\,d\phi\,d\theta = \pi.$$

By symmetry, the center of mass is clearly on the z-axis, so we only need to find the z-coordinate of the center of mass. The moment around the x-y plane is

$$\int_0^{2\pi}\int_0^{\pi/2} z\cos\phi\sin\phi\,d\phi\,d\theta = \int_0^{2\pi}\int_0^{\pi/2} \cos^2\phi\sin\phi\,d\phi\,d\theta = \frac{2\pi}{3},$$

so the center of mass is at $(0,0,2/3)$. ♣

Now suppose that $\mathbf{f}$ is a vector field; imagine that it represents the velocity of some fluid at each point in space. We would like to measure how much fluid is passing through a surface D, the **flux** across D. As usual, we imagine computing the flux across a very small section of the surface, with area dS, and then adding up all such small fluxes over D with an integral. Suppose that vector $\mathbf{N}$ is a unit normal to the surface at a point; $\mathbf{f}\cdot\mathbf{N}$ is the scalar projection of $\mathbf{f}$ onto the direction of $\mathbf{N}$, so it measures how fast the fluid is moving across the surface. In one unit of time the fluid moving across the surface will fill a volume of

$\mathbf{f}\cdot\mathbf{N}\,dS$, which is therefore the rate at which the fluid is moving across a small patch of the surface. Thus, the total flux across D is

$$\iint_D \mathbf{f}\cdot\mathbf{N}\,dS = \iint_D \mathbf{f}\cdot d\mathbf{S},$$

defining $d\mathbf{S} = \mathbf{N}\,dS$.

As usual, certain conditions must be met for this to work out; chief among them is the nature of the surface. As we integrate over the surface, we must choose the normal vectors $\mathbf{N}$ in such a way that they point "the same way" through the surface. For example, if the surface is roughly horizontal in orientation, we might want to measure the flux in the "upwards" direction, or if the surface is closed, like a sphere, we might want to measure the flux "outwards" across the surface. In the first case we would choose $\mathbf{N}$ to have positive z component, in the second we would make sure that $\mathbf{N}$ points away from the origin. Unfortunately, there are surfaces that are not **orientable**: they have only one side, so that it is not possible to choose the normal vectors to point in the "same way" through the surface. The most famous such surface is the Möbius strip shown in Figure 16.7. It is quite easy to make such a strip with a piece of paper and some tape. If you have never done this, it is quite instructive; in particular, you should draw a line down the center of the strip until you return to your starting point. No matter how unit normal vectors are assigned to the points of the Möbius strip, there will be normal vectors very close to each other pointing in opposite directions.

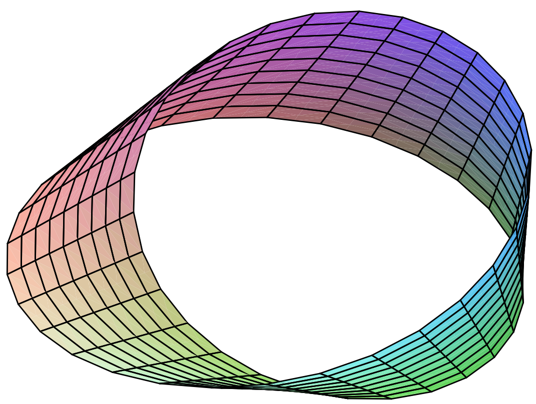


Figure 16.7: A Möbius strip.

Assuming that the quantities involved are well behaved, however, the flux of the vector field across the surface $\mathbf{r}(u,v)$ is

$$\iint_D \mathbf{f}\cdot\mathbf{N}\,dS = \iint_D \mathbf{f}\cdot\frac{\mathbf{r}_u\times\mathbf{r}_v}{|\mathbf{r}_u\times\mathbf{r}_v|}|\mathbf{r}_u\times\mathbf{r}_v|\,dA = \iint_D \mathbf{f}\cdot(\mathbf{r}_u\times\mathbf{r}_v)\,dA.$$

In practice, we may have to use $\mathbf{r}_v\times\mathbf{r}_u$ or even something a bit more complicated to make sure that the normal vector points in the desired direction.

Example 16.22: Flux

Compute the flux of $\mathbf{f} = \langle x, y, z^2\rangle$ *across the cone* $z = \sqrt{x^2+y^2}$, $0 \le z \le 2$, *in the downward direction.*

Solution. We write the cone as a vector function: $\mathbf{r} = \langle v\cos u, v\sin u, v\rangle$, $0 \le u \le 2\pi$ and $0 \le v \le 2$. Then $\mathbf{r}_u = \langle -v\sin u, v\cos u, 0\rangle$ and $\mathbf{r}_v = \langle \cos u, \sin u, 1\rangle$ and $\mathbf{r}_u \times \mathbf{r}_v = \langle v\cos u, v\sin u, -v\rangle$. The third coordinate $-v$ is negative, which is exactly what we desire, that is, the normal vector points down through the surface. Then

$$\begin{aligned}\int_0^{2\pi}\int_0^2 \langle x, y, z^2\rangle \cdot \langle v\cos u, v\sin u, -v\rangle\,dv\,du &= \int_0^{2\pi}\int_0^2 xv\cos u + yv\sin u - z^2 v\,dv\,du\\ &= \int_0^{2\pi}\int_0^2 v^2\cos^2 u + v^2\sin^2 u - v^3\,dv\,du\\ &= \int_0^{2\pi}\int_0^2 v^2 - v^3\,dv\,du = -\frac{8\pi}{3}.\end{aligned}$$

♣

Exercises for 16.7

Exercise 16.7.1 *Find the center of mass of an object that occupies the upper hemisphere of $x^2+y^2+z^2=1$ and has density x^2+y^2.*

Exercise 16.7.2 *Find the center of mass of an object that occupies the surface $z = xy$, $0 \le x \le 1$, $0 \le y \le 1$ and has density $\sqrt{1+x^2+y^2}$.*

Exercise 16.7.3 *Find the center of mass of an object that occupies the surface $z = \sqrt{x^2+y^2}$, $1 \le z \le 4$ and has density $x^2 z$.*

Exercise 16.7.4 *Find the centroid of the surface of a right circular cone of height h and base radius r, not including the base.*

Exercise 16.7.5 *Evaluate $\iint_D \langle 2, -3, 4\rangle \cdot \mathbf{N}\,dS$, where D is given by $z = x^2+y^2$, $-1 \le x \le 1$, $-1 \le y \le 1$, oriented up.*

Exercise 16.7.6 *Evaluate $\iint_D \langle x, y, 3\rangle \cdot \mathbf{N}\,dS$, where D is given by $z = 3x-5y$, $1 \le x \le 2$, $0 \le y \le 2$, oriented up.*

Exercise 16.7.7 *Evaluate $\iint_D \langle x, y, -2\rangle \cdot \mathbf{N}\,dS$, where D is given by $z = 1-x^2-y^2$, $x^2+y^2 \le 1$, oriented up.*

Exercise 16.7.8 *Evaluate $\iint_D \langle e^x, e^y, z\rangle \cdot \mathbf{N}\,dS$, where D is given by $z = xy$, $0 \le x \le 1$, $-x \le y \le x$, oriented up.*

Exercise 16.7.9 *Evaluate $\iint_D \langle xz, yz, z\rangle \cdot \mathbf{N}\,dS$, where D is given by $z = a^2-x^2-y^2$, $x^2+y^2 \le b^2$, oriented up.*

Exercise 16.7.10 *A fluid has density 870 kg/m^3 and flows with velocity $\mathbf{v} = \langle z, y^2, x^2 \rangle$, where distances are in meters and the components of $\mathbf{v}$ are in meters per second. Find the rate of flow outward through the portion of the cylinder $x^2 + y^2 = 4$, $0 \le z \le 1$ for which $y > 0$.*

Exercise 16.7.11 *Gauss's Law says that the net charge, Q, enclosed by a closed surface, S, is*

$$Q = \varepsilon_0 \iint \boldsymbol{e} \cdot \boldsymbol{N}\, dS$$

where $\boldsymbol{e}$ is an electric field and ε_0 (the permittivity of free space) is a known constant; $\boldsymbol{N}$ is oriented outward. Use Gauss's Law to find the charge contained in the cube with vertices $(\pm 1, \pm 1, \pm 1)$ if the electric field is $\boldsymbol{e} = \langle x, y, z \rangle$.

16.8 Stokes' Theorem

Recall that one version of Green's Theorem (see Equation 16.2) is

$$\int_{\partial D} \mathbf{f} \cdot d\mathbf{r} = \iint_D (\nabla \times \mathbf{f}) \cdot \mathbf{k}\, dA.$$

Here D is a region in the x-y plane and $\mathbf{k}$ is a unit normal to D at every point. If D is instead an orientable surface in space, there is an obvious way to alter this equation, and it turns out still to be true:

Theorem 16.23: Stokes' Theorem

Provided that the quantities involved are sufficiently nice, and in particular if D is orientable,

$$\int_{\partial D} \mathbf{f} \cdot d\mathbf{r} = \iint_D (\nabla \times \mathbf{f}) \cdot \mathbf{N}\, dS,$$

if ∂D is oriented counter-clockwise relative to $\mathbf{N}$.

The proof of Stokes' Theorem will follow a discussion and several examples of the Theorem in use.

Note how little has changed: $\mathbf{k}$ becomes $\mathbf{N}$, a unit normal to the surface, and dA becomes dS, since this is now a general surface integral. The phrase "counter-clockwise relative to $\mathbf{N}$" means that if we take the direction of $\mathbf{N}$ to be "up", then we go around the boundary counter-clockwise when viewed from "above".

Example 16.24:

Let $\boldsymbol{f} = \langle y^2 z, x^2 z, x y^2 \rangle$ and the surface D be $x = \sqrt{1 - y^2 - z^2}$, oriented in the positive x direction. It quickly becomes apparent that the surface integral in Stokes' Theorem is intractable, so compute the line integral.

Solution. The boundary of D is the unit circle in the y-z plane, $\mathbf{r} = \langle 0, \cos u, \sin u \rangle$, $0 \le u \le 2\pi$. The integral is

$$\int_0^{2\pi} \langle y^2 z, x^2 z, x y^2 \rangle \cdot \langle 0, -\sin u, \cos u \rangle\, du = \int_0^{2\pi} 0\, du = 0,$$

because $x = 0$. ♣

An interesting consequence of Stokes' Theorem is that if D and E are two orientable surfaces with the same boundary, then

$$\iint_D (\nabla \times \mathbf{f}) \cdot \mathbf{N}\, dS = \int_{\partial D} \mathbf{f} \cdot d\mathbf{r} = \int_{\partial E} \mathbf{f} \cdot d\mathbf{r} = \iint_E (\nabla \times \mathbf{f}) \cdot \mathbf{N}\, dS.$$

Sometimes both of the integrals

$$\iint_D (\nabla \times \mathbf{f}) \cdot \mathbf{N}\, dS \qquad \text{and} \qquad \int_{\partial D} \mathbf{f} \cdot d\mathbf{r}$$

are difficult, but you may be able to find a second surface E so that

$$\iint_E (\nabla \times \mathbf{f}) \cdot \mathbf{N}\, dS$$

has the same value but is easier to compute.

In the previous example, the line integral was easy to compute. But we might also notice that another surface E with the same boundary is the flat disk $y^2 + z^2 \leq 1$.

Example 16.25:

Let $\mathbf{f} = \langle y^2 z, x^2 z, xy^2 \rangle$ and the surface E be $y^2 + z^2 \leq 1$. Compute the surface integral.

Solution. The unit normal $\mathbf{N}$ for this surface is simply $\mathbf{i} = \langle 1, 0, 0 \rangle$. We compute the curl:

$$\nabla \times \mathbf{f} = \langle 2xy - x^2, 0, 2xz - 2yz \rangle.$$

Since $x = 0$ everywhere on the surface,

$$(\nabla \times \mathbf{f}) \cdot \mathbf{N} = \langle 0, 0, 2xz - 2yz \rangle \cdot \langle 1, 0, 0 \rangle = 0,$$

so the surface integral is

$$\iint_E 0\, dS = 0,$$

as before. In this case, of course, it is still somewhat easier to compute the line integral, avoiding $\nabla \times \mathbf{f}$ entirely. ♣

Now let's look at the proof of Stokes' Theorem.

Proof. We can prove here a special case of Stokes' Theorem, which perhaps not too surprisingly uses Green's Theorem.

Suppose the surface D of interest can be expressed in the form $z = g(x, y)$, and let $\mathbf{f} = \langle f_1, f_2, f_3 \rangle$. Using the vector function $\mathbf{r} = \langle x, y, g(x, y) \rangle$ for the surface we get the surface integral

$$\begin{aligned}
\iint_D \nabla \times \mathbf{f} \cdot d\mathbf{S} &= \iint_E \langle \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}, \frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x}, \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \rangle \cdot \langle -g_x, -g_y, 1 \rangle\, dA \\
&= \iint_E -\frac{\partial f_3}{\partial x} g_x + \frac{\partial f_2}{\partial z} g_x - \frac{\partial f_1}{\partial z} g_y + \frac{\partial f_3}{\partial x} g_y + \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\, dA.
\end{aligned}$$

Here E is the region in the x-y plane directly below the surface D.

For the line integral, we need a vector function for ∂D. If $\langle x(t), y(t)\rangle$ is a vector function for ∂E then we may use $\mathbf{r}(t) = \langle x(t), y(t), g(x(t), y(t))\rangle$ to represent ∂D. Then

$$\int_{\partial D} \mathbf{f}\cdot d\mathbf{r} = \int_a^b f_1\frac{dx}{dt} + f_2\frac{dy}{dt} + f_3\frac{dz}{dt}\,dt = \int_a^b f_1\frac{dx}{dt} + f_2\frac{dy}{dt} + f_2\left(\frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\right)dt.$$

using the chain rule for dz/dt. Now we continue to manipulate this:

$$\begin{aligned}
&\int_a^b f_1\frac{dx}{dt} + f_2\frac{dy}{dt} + f_3\left(\frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\right)dt\\
&\qquad = \int_a^b \left[\left(f_1 + f_3\frac{\partial z}{\partial x}\right)\frac{dx}{dt} + \left(f_2 + f_3\frac{\partial z}{\partial y}\right)\frac{dy}{dt}\right]dt\\
&\qquad = \int_{\partial E} \left(f_1 + f_3\frac{\partial z}{\partial x}\right)dx + \left(f_2 + f_3\frac{\partial z}{\partial y}\right)dy,
\end{aligned}$$

which now looks just like the line integral of Green's Theorem, except that the functions f_1 and f_2 of Green's Theorem have been replaced by the more complicated $f_1 + f_3(\partial z/\partial x)$ and $f_2 + f_3(\partial z/\partial y)$. We can apply Green's Theorem to get

$$\int_{\partial E} \left(f_1 + f_3\frac{\partial z}{\partial x}\right)dx + \left(f_2 + f_3\frac{\partial z}{\partial y}\right)dy = \iint_E \frac{\partial}{\partial x}\left(f_2 + f_3\frac{\partial z}{\partial y}\right) - \frac{\partial}{\partial y}\left(f_1 + f_3\frac{\partial z}{\partial x}\right)dA.$$

Now we can use the chain rule again to evaluate the derivatives inside this integral, and it becomes

$$\begin{aligned}
&\iint_E \frac{\partial f_2}{\partial x} + \frac{\partial f_2}{\partial z}g_x + \frac{\partial f_3}{\partial x}g_y + \frac{\partial f_3}{\partial z}g_xg_y + f_3g_{yx} - \left(\frac{\partial f_1}{\partial y} + \frac{\partial f_1}{\partial z}g_y + \frac{\partial f_3}{\partial y}g_x + \frac{\partial f_3}{\partial z}g_yg_x + f_3g_{xy}\right)dA\\
&\quad = \iint_E \frac{\partial f_2}{\partial x} + \frac{\partial f_2}{\partial z}g_x + \frac{\partial f_3}{\partial x}g_y - \frac{\partial f_1}{\partial y} - \frac{\partial f_1}{\partial z}g_y - \frac{\partial f_3}{\partial y}g_x\,dA,
\end{aligned}$$

which is the same as the expression we obtained for the surface integral. ♣

Exercises for 16.8

Exercise 16.8.1 *Let $\boldsymbol{f} = \langle z, x, y\rangle$. The plane $z = 2x + 2y - 1$ and the paraboloid $z = x^2 + y^2$ intersect in a closed curve. Stokes' Theorem implies that*

$$\iint_{D_1} (\nabla\times\boldsymbol{f})\cdot\boldsymbol{N}\,dS = \oint_C \boldsymbol{f}\cdot d\boldsymbol{r} = \iint_{D_2} (\nabla\times\boldsymbol{f})\cdot\boldsymbol{N}\,dS,$$

where the line integral is computed over the intersection C of the plane and the paraboloid, and the two surface integrals are computed over the portions of the two surfaces that have boundary C (provided, of course, that the orientations all match). Compute all three integrals.

Exercise 16.8.2 *Let D be the portion of $z = 1 - x^2 - y^2$ above the x-y plane, oriented up, and let $\boldsymbol{f} = \langle xy^2, -x^2y, xyz\rangle$. Compute $\iint_D (\nabla\times\boldsymbol{f})\cdot\boldsymbol{N}\,dS$.*

Exercise 16.8.3 *Let D be the portion of* $z = 2x+5y$ *inside* $x^2+y^2 = 1$, *oriented up, and let* $\boldsymbol{f} = \langle y, z, -x\rangle$. *Compute* $\displaystyle\int_{\partial D} \boldsymbol{f}\cdot d\boldsymbol{r}$.

Exercise 16.8.4 *Compute* $\displaystyle\oint_C x^2 z\,dx + 3x\,dy - y^3\,dz$, *where C is the unit circle* $x^2+y^2 = 1$ *oriented counter-clockwise.*

Exercise 16.8.5 *Let D be the portion of* $z = px+qy+r$ *over a region in the x-y plane that has area A, oriented up, and let* $\boldsymbol{f} = \langle ax+by+cz, ax+by+cz, ax+by+cz\rangle$. *Compute* $\displaystyle\int_{\partial D} \boldsymbol{f}\cdot d\boldsymbol{r}$.

Exercise 16.8.6 *Let D be any surface and let* $\boldsymbol{f} = \langle f_1(x), f_2(y), f_3(z)\rangle$ *(*f_1 *depends only on x,* f_2 *only on y, and* f_3 *only on z). Show that* $\displaystyle\int_{\partial D} \boldsymbol{f}\cdot d\boldsymbol{r} = 0$.

Exercise 16.8.7 *Show that* $\displaystyle\int_C f\nabla g + g\nabla f\cdot d\boldsymbol{r} = 0$, *where* $\boldsymbol{r}$ *describes a closed curve C to which Stokes' Theorem applies.*

Selected Exercise Answers

1.1.1 (a) $\dfrac{x}{y}$

(b) $\sqrt{x}y$

(c) $\dfrac{2}{\sqrt[3]{x}}$

1.1.2 $a = 2$, $b = -\frac{5}{3}$, $c = \frac{3}{2}$.

1.1.3 $x = -4$ and $x = 6$.

1.1.5 (a) $(5/3, \infty)$

(b) $[1/7, 2/7]$

(c) $(-\infty, -3) \cup (-2, 1]$

(d) $(-\infty, \infty)$

(e) No solution

(f) $(-\infty, 1) \cup (1, \infty)$

(g) $(-2, 0) \cup (2, \infty)$

(h) $[4, \infty) \cup \{0\}$

(i) $(0, \frac{1}{2})$

(j) $(-2, -1] \cup (1, 4]$

1.1.6 $x = -\frac{1}{2}$ and $x = -\frac{1}{6}$.

1.1.7 (a) $(-\infty, -2] \cup [2, \infty)$

(b) $[2, 4]$

(c) $(-\infty, -9/2] \cup [-1/2, \infty)$

(d) $(4, \infty)$

(e) $(-\infty, \infty)$

(f) $(-9, -6) \cup (4, 7)$

1.2.1 (a) $(2/3)x + (1/3)$

(b) $y = -2x$

(c) $y = (-2/3)x + (1/3)$

(d) $y = -x/3 + 17/3$

(e) $y = -1/2x + 5/2$

1.2.2 (a) $y = 2x + 2$, 2, -1

(b) $y = -x + 6$, 6, 6

(c) $y = x/2 + 1/2$, $1/2$, -1

(d) $y = 3/2$, y-intercept: $3/2$, no x-intercept

(e) $y = (-2/3)x - 2$, -2, -3

1.2.3 Yes, the lines are parallel as they have the same slope of $-1/2$

1.2.4 $y = 0$, $y = -2x + 2$, $y = 2x + 2$

1.2.5 $y = (9/5)x + 32$, $(-40, -40)$

1.2.6 $y = 0.15x + 10$

1.2.7 $0.03x + 1.2$

1.2.8 (a) $P = -0.0001x + 2$
(b) $x = -10000P + 20000$

1.2.9 $(2/25)x - (16/5)$

1.2.10 (a) 2

(b) $\sqrt{2}$

(c) $\sqrt{2}$

1.2.12 (a) $x^2 + y^2 = 9$

(b) $(x-5)^2 + (y-6)^2 = 9$

(c) $(x+5)^2 + (y+6)^2 = 9$

1.2.14 (a) circle

(b) ellipse

(c) horizontal parabola

1.2.15 $(x + 2/7)^2 + (y - 41/7)^2 = 1300/49$

1.3.1 $2n\pi - \pi/2$, any integer n

1.3.2 $n\pi \pm \pi/6$, any integer n

1.3.4 $-\frac{5}{4}$

1.3.5 $\sin\theta = -x/\sqrt{x^2+1}$, $\cos\theta = -1/\sqrt{x^2+1}$.

1.3.6 $-\frac{2\pi}{7}$ is the unique answer.

1.3.7 $(\sqrt{2}+\sqrt{6})/4$

1.3.8 $-(1+\sqrt{3})/(1-\sqrt{3}) = 2+\sqrt{3}$

1.3.11 $t = \pi/2$

1.4.1 (a) $\dfrac{\sqrt{2}}{2}$

(b) $3(\sqrt{x+h+1}+\sqrt{x+1})$

1.4.2 (a) $-13/5$

(b) $-1/2, 3$

(c) $(1\pm\sqrt{13})/2$

(d) No real solutions

(e) $\sqrt{3}$

1.4.3 Counter-examples may vary.

(a) $x = 3$

(b) $x = h = 1$

(c) $x = y = 1$

1.4.4 $x+3y-13 = 0$, or equivalents such as $y = -\frac{1}{3}x+\frac{13}{3}$

1.4.5 $(-\infty, 0] \cup (\frac{1}{3}, 3]$

1.4.6 It is impossible for both $x-2$ and $1-x$ to be non-negative for the same real number x.

1.4.7 $6x+3h$

1.4.8 $-1/[(2x+2h-1)(2x-1)]$

1.4.9 -3

1.4.10 $\pi/6, 5\pi/6$

1.4.11 $2\pi/5$

1.4.12 It is equal to 2 for all x larger than 4.

1.4.13 $(x+2)^2+(y-3)^2=25$.

1.4.14 Centre is $(-3,2)$ and radius is 2.

1.4.15 y could be any real number greater than or equal to 6.

1.4.16 $x^6y^4/(36z^8)$

1.4.17 $x=(y-2)/(3+4y)$

1.4.18 $Q(x)=x+1$, $R=-7$

2.1.1 (a) $\{x \mid x \in \mathbb{R}\}$, i.e., all x

(b) $\{x \mid x \geq 3/2\}$

(c) $\{x \mid x \neq -1\}$

(d) $\{x \mid x \neq 1 \text{ and } x \neq -1\}$

(e) $\{x \mid x < 0\}$

(f) $\{x \mid x \in \mathbb{R}\}$, i.e., all x

(g) $\{x \mid h-r \leq x \leq h+r\}$

(h) $\{x \mid x \geq 0\}$

(i) $\{x \mid -1 \leq x \leq 1\}$

(j) $\{x \mid x \geq 1\}$

(k) $\{x \mid -1/3 < x < 1/3\}$

(l) $\{x \mid x \geq 0 \text{ and } x \neq 1\}$

(m) $\{x \mid x \geq 0 \text{ and } x \neq 1\}$

2.1.2 $A=x(500-2x)$, $\{x \mid 0 \leq x \leq 250\}$

2.1.3 $V=r(50-\pi r^2)$, $\{r \mid 0 < r \leq \sqrt{50/\pi}\}$

2.1.4 $A = 2\pi r^2 + 2000/r$, $\{r \mid 0 < r < \infty\}$

2.2.3 $\{x \mid x \geq 3\}$, $\{x \mid x \geq 0\}$

2.3.1 $y = 2^x$

2.3.2 $y = 7$

2.3.3 $y = 2$

2.3.4 $x \neq 0$

2.6.1 (a) $\pi/3$

(b) $3\pi/4$

2.6.2 (a) $\pi/4$

(b) $-\pi/3$

(c) $1/3$

(d) $-3/4$

2.6.3 $\sqrt{1-x^2}/x$ with domain $[-1,0) \cup (0,1]$.

2.8.1 (d)

2.8.2 $3/[5(x+3)]$

2.8.3 $6+h$

2.8.4 (a) $[2,3) \cup (3,\infty)$

(b) $(-\infty,-3) \cup (-3,3) \cup (3,\infty)$

2.8.5 $\{x : x \neq 0\}$

2.8.6 $g(x) = (5x+26)/3$

2.8.7 (c)

2.8.8 $f^{-1}(x) = f(x) = \ln\left(\dfrac{e^x}{e^x - 1}\right)$ and its domain is $(0,\infty)$.

2.8.9 (a) $2 - \ln 3$

(b) 1,3

(c) $(e^2-1)^2$

(d) 3

2.8.10 $-\pi$

2.8.11 $2\pi/5$

2.8.12 1

3.3.1

(a) 8
(b) 6
(c) dne
(d) -2
(e) -1
(f) 8
(g) 7
(h) 6
(i) 3
(j) $-3/2$
(k) 6
(l) 2

3.4.1 (a) 7
(b) 5
(c) 0
(d) undefined
(e) $1/6$
(f) 0
(g) 3
(h) 172
(i) 0
(j) 2
(k) does not exist
(l) $\sqrt{2}$
(m) $3a^2$
(n) 512

3.4.2 $L=0$ and $M=1$. No.

3.5.1 (a) 1
(b) 1
(c) $-\infty$
(d) $1/3$
(e) 0
(f) ∞
(g) ∞
(h) $2/7$
(i) 2
(j) $-\infty$
(k) ∞
(l) 0
(m) $1/2$
(n) 5
(o) $2\sqrt{2}$
(p) $3/2$
(q) ∞
(r) does not exist

3.5.2 $y=1$ and $y=-1$

3.5.3 $x=0$ and $x=2$.

3.5.5 $y=x+4$

3.5.6 (a) $-\infty$
(b) $\pi/2$
(c) 0
(d) ∞
(e) -5

(f) 1/3

(g) 0

3.6.1 (a) 5

(b) $7/2$

(c) $3/4$

(d) 1

(e) $-\sqrt{2}/2$

3.6.2 7

3.6.3 2

3.6.5 3

4.1.1 -5, -2.47106145, -2.4067927, -2.400676, -2.4

4.1.2 $-4/3$, $-24/7$, $7/24$, $3/4$

4.1.3 -0.107526881, -0.11074197, -0.1110741, $\dfrac{-1}{3(3+\Delta x)} \to \dfrac{-1}{9}$

4.1.4 $\dfrac{3+3\Delta x+\Delta x^2}{1+\Delta x} \to 3$

4.1.5 3.31, 3.003001, 3.0000,
$3+3\Delta x+\Delta x^2 \to 3$

4.1.6 m

4.1.9 10, $25/2$, 20, 15, 25, 35.

4.1.10 5, 4.1, 4.01, 4.001, $4+\Delta t \to 4$

4.1.11 -10.29, -9.849, -9.8049,
$-9.8-4.9\Delta t \to -9.8$

4.2.1 (a) $-x/\sqrt{169-x^2}$

(b) $-9.8t$

(c) $2x+1/x^2$

(d) $2ax+b$

(e) $3x^2$

(f) $-2/(2x+1)^{3/2}$

(g) $5/(t+2)^2$

4.2.4 $y=-13x+17$

4.2.5 -8

4.3.1 (a) $100x^{99}$

(b) $-100x^{-101}$

(c) $-5x^{-6}$

(d) $\pi x^{\pi-1}$

(e) $(3/4)x^{-1/4}$

(f) $-(9/7)x^{-16/7}$

(g) $15x^2+24x$

(h) $-20x^4+6x+10/x^3$

(i) $-30x+25$

(j) $3x^2+6x-1$

(k) $-\dfrac{x^2+2x+5}{(x^2+2x-3)^2}$

(l) $3x^2(x^3-5x+10)+x^3(3x^2-5)$

(m) $x^4(7x^2+30x-15)$

(n) $\dfrac{-3x^2-20x+15}{x^6}$

(o) $-\dfrac{3x(5x+8)}{(5x^3+12x^2-15)^2}$

4.3.2 $y=13x/4+5$

4.3.3 $y=24x-48-\pi^3$

4.3.4 $-49t/5+5$, $-49/5$

4.3.6 $\sum_{k=1}^{n} ka_kx^{k-1}$

4.3.7 $x^3/16-3x/4+4$

4.3.10 $f'=4(2x-3)$, $y=4x-7$

4.3.12 $\dfrac{3x^2}{x^3-5x+10}-\dfrac{x^3(3x^2-5)}{(x^3-5x+10)^2}$

4.3.13 $\dfrac{2x+5}{x^5-6x^3+3x^2-7x+1}-\dfrac{(x^2+5x-3)(5x^4-18x^2+6x-7)}{(x^5-6x^3+3x^2-7x+1)^2}$

4.3.14 $\dfrac{x-1250}{2(x-625)^{3/2}}$

4.3.15 $\dfrac{200-39x}{2x^{21}\sqrt{x-5}}$

4.3.16 $y=17x/4-41/4$

4.3.17 $y = 11x/16 - 15/16$

4.3.18 $13/18$

4.4.2 $\pi/6 + 2n\pi$, $5\pi/6 + 2n\pi$, any integer n

4.5.1 $4x^3 - 9x^2 + x + 7$

4.5.2 $3x^2 - 4x + 2/\sqrt{x}$

4.5.3 $6(x^2+1)^2 x$

4.5.4 $\sqrt{169 - x^2} - x^2/\sqrt{169 - x^2}$

4.5.5 $(2x-4)\sqrt{25-x^2} -$
$(x^2 - 4x + 5)x/\sqrt{25 - x^2}$

4.5.6 $-x/\sqrt{r^2 - x^2}$

4.5.7 $2x^3/\sqrt{1+x^4}$

4.5.8 $\dfrac{1}{4\sqrt{x}(5-\sqrt{x})^{3/2}}$

4.5.9 $6 + 18x$

4.5.10 $\dfrac{2x+1}{1-x} + \dfrac{x^2+x+1}{(1-x)^2}$

4.5.11 $-1/\sqrt{25-x^2} - \sqrt{25-x^2}/x^2$

4.5.12 $\dfrac{1}{2}\left(\dfrac{-169}{x^2} - 1\right) \bigg/ \sqrt{\dfrac{169}{x} - x}$

4.5.13 $\dfrac{3x^2 - 2x + 1/x^2}{2\sqrt{x^3 - x^2 - (1/x)}}$

4.5.14 $\dfrac{300x}{(100 - x^2)^{5/2}}$

4.5.15 $\dfrac{1+3x^2}{3(x+x^3)^{2/3}}$

4.5.16 $\left(4x(x^2+1)+\dfrac{4x^3+4x}{2\sqrt{1+(x^2+1)^2}}\right)\Big/$
$2\sqrt{(x^2+1)^2+\sqrt{1+(x^2+1)^2}}$

4.5.17 $5(x+8)^4$

4.5.18 $-3(4-x)^2$

4.5.19 $6x(x^2+5)^2$

4.5.20 $-12x(6-2x^2)^2$

4.5.21 $24x^2(1-4x^3)^{-3}$

4.5.22 $5+5/x^2$

4.5.23 $-8(4x-1)(2x^2-x+3)^{-3}$

4.5.24 $1/(x+1)^2$

4.5.25 $3(8x-2)/(4x^2-2x+1)^2$

4.5.26 $-3x^2+5x-1$

4.5.27 $6x(2x-4)^3+6(3x^2+1)(2x-4)^2$

4.5.28 $-2/(x-1)^2$

4.5.29 $4x/(x^2+1)^2$

4.5.30 $(x^2-6x+7)/(x-3)^2$

4.5.31 $-5/(3x-4)^2$

4.5.32 $60x^4+72x^3+18x^2+18x-6$

4.5.33 $(5-4x)/((2x+1)^2(x-3)^2)$

4.5.34 $1/(2(2+3x)^2)$

4.5.35 $56x^6+72x^5+110x^4+100x^3+60x^2+28x+6$

4.5.36 $y=23x/96-29/96$

4.5.37 $y=3-2x/3$

4.5.38 $y = 13x/2 - 23/2$

4.5.39 $y = 2x - 11$

4.5.40 $y = \dfrac{20 + 2\sqrt{5}}{5\sqrt{4+\sqrt{5}}}x + \dfrac{3\sqrt{5}}{5\sqrt{4+\sqrt{5}}}$

4.5.41 $(f(g(1)))' = 20$

4.5.42 $g'(x) = 2x(f(x^2) + x^2 f'(x^2))$

4.6.1 $2\ln(3)x3^{x^2}$

4.6.2 $\dfrac{\cos x - \sin x}{e^x}$

4.6.3 $2e^{2x}$

4.6.4 $e^x \cos(e^x)$

4.6.5 $\cos(x)e^{\sin x}$

4.6.6 $x^{\sin x}\left(\cos x \ln x + \dfrac{\sin x}{x}\right)$

4.6.7 $3x^2e^x + x^3e^x$

4.6.8 $1 + 2^x \ln(2)$

4.6.9 $-2x\ln(3)(1/3)^{x^2}$

4.6.10 $e^{4x}(4x-1)/x^2$

4.6.11 $(3x^2+3)/(x^3+3x)$

4.6.12 $-\tan(x)$

4.6.13 $(1-\ln(x^2))/(x^2\sqrt{\ln(x^2)})$

4.6.14 $\sec(x)$

4.6.15 $x^{\cos(x)}(\cos(x)/x - \cos(x)\ln(x))$

4.6.19 e

4.7.1 (a) x/y

(b) $-(2x+y)/(x+2y)$

(c) $(2xy-3x^2-y^2)/(2xy-3y^2-x^2)$

(d) $\sin(x)\sin(y)/(\cos(x)\cos(y))$

(e) $-\sqrt{y}/\sqrt{x}$

(f) $(y\sec^2(x/y)-y^2)/(x\sec^2(x/y)+y^2)$

(g) $(y-\cos(x+y))/(\cos(x+y)-x)$

(h) $-y^2/x^2$

4.7.2 1

4.7.3 $y=2x\pm 6$

4.7.4 $y=x/2\pm 3$

4.7.6 $(\sqrt{3},2\sqrt{3})$, $(-\sqrt{3},-2\sqrt{3})$, $(2\sqrt{3},\sqrt{3})$, $(-2\sqrt{3},-\sqrt{3})$

4.7.7 $y=7x/\sqrt{3}-8/\sqrt{3}$

4.7.8 $y=(-y_1^{1/3}x+y_1^{1/3}x_1+x_1^{1/3}y_1)/x_1^{1/3}$

4.7.9 $(y-y_1)/(x-x_1)=(2x_1^3+2x_1y_1^2-x_1)/(2y_1^3+2y_1x_1^2+y_1)$

4.8.1 1

4.9.1 (a) $4(2x+3)$

(b) $\frac{3}{2}x^{1/2}$

4.9.2 3

4.9.3 (a) $28x^3-\dfrac{1}{3\pi x^{4/3}}$

(b) $-\dfrac{1}{\sqrt{x}(1+\sqrt{x})^2}$

(c) $f'(x)=\begin{cases} -2 & \text{if } x<-2 \\ 0 & \text{if } -2<x<1 \\ 2 & \text{if } x>1 \end{cases}$

(d) $2x\sin x\cos x+x^2\cos^2 x-x^2\sin^2 x$

(e) $\dfrac{(\sin x + x\cos x)(1+\sin x) - x\sin x\cos x}{(1+\sin x)^2}$

(f) $-\dfrac{3}{4x^{3/2}}\left(2+\dfrac{3}{\sqrt{x}}\right)^{-1/2}$

(g) $\frac{1}{3}(x^4+x^2+1)^{-2/3}(4x^3+2x) - \dfrac{5(3x^2-1)}{(x^3-x+4)^6}$

(h) $3\sin^2 x\cos x - 3x^2\cos(x^3)$

(i) $4\sec^4 x\tan x + 4\tan^3 x\sec^2 x$

(j) $\dfrac{4}{(1+x)^2}\cos\left(\dfrac{1-x}{1+x}\right)\sin\left(\dfrac{1-x}{1+x}\right)$

(k) $(2x+2\sec^2 x\tan x)\sec^2(\sin(x^2+\sec^2 x))\cos(x^2+\sec^2 x)$

(l) $-\dfrac{\pi\cos\frac{\pi}{x}}{x^2(2+\sin\frac{\pi}{x})^2}$

4.9.4 (a) $3e^{3x} - e^{-x}$

(b) $2e^{2x}\cos 3x - 3e^{2x}\sin 3x$

(c) $(1+e^x)\sec^2(x+e^x)$

(d) $2e^x/(e^x+2)^2$

(e) $\dfrac{\cos x}{2+\sin x} - \dfrac{\cos(2+\ln x)}{x}$

(f) $e^{x^\pi}\cdot\pi x^{\pi-1} + \pi^e x^{\pi^e-1} + \pi^{e^x}\ln\pi\cdot e^x$

(g) $\log_a b + (\log_a b)x^{(\log_a b)-1}$

(h) $(x^2+1)^{x^3+1}\left(3x^2\ln(x^2+1) + \dfrac{2x(x^3+1)}{x^2+1}\right)$

(i) $\dfrac{(x^2+e^x)^{1/\ln x}}{(\ln x)^2}\left(\dfrac{2x+e^x}{x^2+e^x}\ln x - \dfrac{x^2+e^x}{x}\right)$

(j) $\dfrac{x\sqrt{x^2+x+1}}{(2+\sin x)^4(3x+5)^7}\left(\dfrac{1}{x}+\dfrac{2x+1}{2(x^2+x+1)} - \dfrac{4\cos x}{2+\sin x} - \dfrac{21}{3x+5}\right)$

4.9.5 (a) $-(2x+y)/(x+2y)$

(b) $\dfrac{x-(2x^2+2y^2-x)(4x-1)}{4y(2x^2+2y^2-x)-y}$

(c) $-\dfrac{\sin x+2x\sin y}{x^2\cos y+3y^2}$

(d) $\dfrac{2x+e^y-e^x}{2-xe^y}$

4.9.6 (a) $\sin^{-1}x+x/\sqrt{1-x^2}$

(b) $\dfrac{\cos^{-1}x+\sin^{-1}x}{(\cos^{-1}x)^2\sqrt{1-x^2}}$

(c) $a/(x^2+a^2)$

(d) $\tan^{-1}x$

5.1.2 $1/(16\pi)$ cm/s

5.1.3 $3/(1000\pi)$ meters/second

5.1.4 $1/4$ m/s

5.1.5 $-6/25$ m/s

5.1.6 80π mi/min

5.1.7 $3\sqrt{5}$ ft/s

5.1.8 $20/(3\pi)$ cm/s

5.1.9 $13/20$ ft/s

5.1.10 $5\sqrt{10}/2$ m/s

5.1.11 $75/64$ m/min

5.1.12 tip: 6 ft/s, length: $5/2$ ft/s

5.1.13 tip: $20/11$ m/s, length: $9/11$ m/s

5.1.14 $380/\sqrt{3}-150\approx 69.4$ mph

5.1.15 $500/\sqrt{3}-200\approx 88.7$ km/hr

5.1.16 $4000/49$ m/s

5.2.1 min at $x=1/2$

5.2.2 min at $x = -1$, max at $x = 1$

5.2.3 max at $x = 2$, min at $x = 4$

5.2.4 min at $x = \pm 1$, max at $x = 0$.

5.2.5 min at $x = 1$

5.2.6 none

5.2.7 none

5.2.8 min at $x = 7\pi/12 + k\pi$, max at $x = -\pi/12 + k\pi$, for integer k.

5.2.9 local min at $x = 49$

5.2.12 one

5.2.16 Absolute maximum $(3,7)$; Absolute minimum $(0,1)$.

5.2.17 Absolute maximum $(3,7)$; Absolute minimum $(0,1)$.

5.2.18 Absolute minimum $(\pi/2,1)$; No absolute maximum.

5.2.19 Absolute minimum $(1,0)$; Absolute maximum $(e^{1/2}, \frac{1}{2e})$.

5.2.20 Absolute minimum $(1,0)$; Absolute maximum $(e^{1/2}, \frac{1}{2e})$.

5.2.21 Absolute minimum $(0,0)$; Absolute maximum $(2, 2e^{1/8})$.

5.2.22 Absolute minimum $(1/2, \frac{2-\pi}{4})$; Absolute maximum $(2, 2-\tan^{-1}(4))$.

5.2.23 Absolute maximum $(1,1/2)$; Absolute minimum $(-1,-1/2)$.

5.3.1 $c = 1/2$

5.3.2 $c = \sqrt{18} - 2$

5.3.6 $x^3/3 + 47x^2/2 - 5x + k$

5.3.7 $\arctan x + k$

5.3.8 $x^4/4 - \ln x + k$

5.3.9 $-\cos(2x)/2 + k$

5.4.1 $L(x) = x$, $f(0.1) \approx L(0.1) = 0.1$

5.4.2 Choose $f(x) = x^3$ and $a = 2$, the closest integer to 1.9. The linearization of f at a is $L(x) = 12(x-2)+8$, and $(1.9)^3 = f(1.9) \approx L(1.9) = 12(1.9-2)+8 = 6.8$.

5.4.4 Choose $a = 7$ since $f(7) = \sqrt[3]{7+1} = \sqrt[3]{8} = 2$ is an integer close to $\sqrt[3]{9}$. The linearization of f at $a = 7$ is $L(x) = 1/12(x-7)+2$. Then $f(8) = \sqrt[3]{8+1} = \sqrt[3]{9} \approx L(8) = 1/12(8-7)+2 = 2.08\bar{3}$. We are over-estimating $\sqrt[3]{9}$ since $L(x) > f(x)$ for all x around $a = 7$.

5.4.5 $\Delta y = 65/16$, $dy = 2$

5.4.6 $\Delta y = \sqrt{11/10} - 1$, $dy = 0.05$

5.4.7 $\Delta y = \sin(\pi/50)$, $dy = \pi/50$

5.4.8 $dV = 8\pi/25$

5.4.9 $T_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}$

(a) $\sin(0.1) \approx T_5(0.1) \approx 0.10016675$

(b) $\sin(0.1) = 0.0998334\ldots$ using a calculator. Our approximation is accurate to $0.10016675 - 0.0998334\ldots = 0.000\bar{3}$.

5.4.11 $T_3(x) = x + x^2 + x^3$. The point $x = 5$ is not close to $x = 0$, and f is not continuous at $x = 1$.

5.4.12 (a) $f^{(n)}(x) = \frac{(-1)^{(n-1)}(n-1)!}{x^n}$

(b) $T_n(x) = \ln(1) + \sum_{i=1}^{n} \frac{\left(\frac{(-1)^{(i-1)}(i-1)!}{1^n}\right)}{i!}(x-1)^i = \sum_{i=1}^{n} \left(\frac{(-1)^{(i-1)}(i-1)!}{i!}\right)(x-1)^i$ since $\ln(1) = 0$ and $1^n = 1$.

5.4.13 Notice that $f(-2) = 19$, $f(0) = -11$, and $f(5) = 19$ and f is a continuous function. By the Intermediate Value Theorem there exists a root in $[-2,0]$ and $[0,5]$. Choose $x_0 = 0$, then $x_4 \approx -0.93242$. Choose $x_0 = 5$, then $x_4 \approx 3.93242$.

5.4.14 (a) $x_4 \approx 1.00022\ldots$

(b) $x = 1$ is the root of f. Our approximation in part (a) was correct to 3 decimal places.

(c) $x_1 = 1$. The root is found in one iteration of Newton's Method.

5.4.15 $\cos(\pi/2) = 0$, so x_1 is undefined.

5.5.1 0

5.5.2 ∞

5.5.3 0

5.5.4 0

5.5.5 $1/6$

5.5.6 $1/16$

5.5.7 $3/2$

5.5.8 $-1/4$

5.5.9 -3

5.5.10 $1/2$

5.5.11 0

5.5.12 -1

5.5.13 $-1/2$

5.5.14 5

5.5.15 1

5.5.16 1

5.5.17 2

5.5.18 1

5.5.19 0

5.5.20 $1/2$

5.5.21 2

5.5.22 0

5.5.23 $1/2$

5.5.24 $-1/2$

5.5.25 2

5.5.26 0

5.5.27 ∞

5.5.28 0

5.5.29 5

5.5.30 $-1/2$

5.6.1 min at $x = 1/2$

5.6.2 min at $x = -1$, max at $x = 1$

5.6.3 max at $x = 2$, min at $x = 4$

5.6.4 min at $x = \pm 1$, max at $x = 0$.

5.6.5 min at $x = 1$

5.6.6 none

5.6.7 none

5.6.8 min at $x = 7\pi/12 + k\pi$, max at $x = -\pi/12 + k\pi$, for integer k.

5.6.9 none

5.6.10 max at $x = 0$, min at $x = \pm 11$

5.6.11 min at $x = -3/2$, neither at $x = 0$

5.6.12 min at $n\pi$, max at $\pi/2 + n\pi$

5.6.13 min at $2n\pi$, max at $(2n+1)\pi$

5.6.14 min at $\pi/2 + 2n\pi$, max at $3\pi/2 + 2n\pi$

5.6.17 min at $x = 1/2$

5.6.18 min at $x = -1$, max at $x = 1$

5.6.19 max at $x = 2$, min at $x = 4$

5.6.20 min at $x = \pm 1$, max at $x = 0$.

5.6.21 min at $x = 1$

5.6.22 none

5.6.23 none

5.6.24 min at $x = 7\pi/12 + n\pi$, max at $x = -\pi/12 + n\pi$, for integer n.

5.6.25 max at $x = 63/64$

5.6.26 max at $x = 7$

5.6.27 max at $-5^{-1/4}$, min at $5^{-1/4}$

5.6.28 none

5.6.29 max at -1, min at 1

5.6.30 min at $2^{-1/3}$

5.6.31 none

5.6.32 min at $n\pi$

5.6.33 max at $n\pi$, min at $\pi/2 + n\pi$

5.6.34 max at $\pi/2 + 2n\pi$, min at $3\pi/2 + 2n\pi$

5.6.35 concave up everywhere

5.6.36 concave up when $x < 0$, concave down when $x > 0$

5.6.37 concave down when $x < 3$, concave up when $x > 3$

5.6.38 concave up when $x < -1/\sqrt{3}$ or $x > 1/\sqrt{3}$, concave down when $-1/\sqrt{3} < x < 1/\sqrt{3}$

5.6.39 concave up when $x < 0$ or $x > 2/3$, concave down when $0 < x < 2/3$

5.6.40 concave up when $x < 0$, concave down when $x > 0$

5.6.41 concave up when $x < -1$ or $x > 1$, concave down when $-1 < x < 0$ or $0 < x < 1$

5.6.42 concave down on $((8n-1)\pi/4, (8n+3)\pi/4)$, concave up on $((8n+3)\pi/4, (8n+7)\pi/4)$, for integer n

5.6.43 concave down everywhere

5.6.44 concave up on $(-\infty,(21-\sqrt{497})/4)$ and $(21+\sqrt{497})/4,\infty)$

5.6.45 concave up on $(0,\infty)$

5.6.46 concave down on $(2n\pi/3,(2n+1)\pi/3)$

5.6.47 concave up on $(0,\infty)$

5.6.48 concave up on $(-\infty,-1)$ and $(0,\infty)$

5.6.49 concave down everywhere

5.6.50 concave up everywhere

5.6.51 concave up on $(\pi/4+n\pi,3\pi/4+n\pi)$

5.6.52 inflection points at $n\pi$, $\pm\arcsin(\sqrt{2/3})+n\pi$

5.6.53 up/incr: $(3,\infty)$, up/decr: $(-\infty,0)$, $(2,3)$, down/decr: $(0,2)$

5.7.1 25×25

5.7.2 $P/4\times P/4$

5.7.3 $w=l=2\cdot 5^{2/3}$, $h=5^{2/3}$, $h/w=1/2$

5.7.4 $\sqrt[3]{100}\times\sqrt[3]{100}\times 2\sqrt[3]{100}$, $h/s=2$

5.7.5 $w=l=2^{1/3}V^{1/3}$, $h=V^{1/3}/2^{2/3}$, $h/w=1/2$

5.7.6 1250 square feet

5.7.7 $l^2/8$ square feet

5.7.8 \$5000

5.7.9 100

5.7.10 r^2

5.7.11 $h/r=2$

5.7.12 $h/r=2$

5.7.13 $r = 5$, $h = 40/\pi$, $h/r = 8/\pi$

5.7.14 $8/\pi$

5.7.15 $4/27$

5.7.16 (a) 2, (b) $7/2$

5.7.17 $\dfrac{\sqrt{3}}{6} \times \dfrac{\sqrt{3}}{6} + \dfrac{1}{2} \times \dfrac{1}{4} - \dfrac{\sqrt{3}}{12}$

5.7.18 (a) $a/6$, (b) $(a + b - \sqrt{a^2 - ab + b^2})/6$

5.7.19 1.5 meters wide by 1.25 meters tall

5.7.20 If $k \le 2/\pi$ the ratio is $(2 - k\pi)/4$; if $k \ge 2/\pi$, the ratio is zero: the window should be semicircular with no rectangular part.

5.7.21 a/b

5.7.22 $1/\sqrt{3} \approx 58\%$

5.7.23 $18 \times 18 \times 36$

5.7.24 $r = 5/(2\pi)^{1/3} \approx 2.7$ cm,
$h = 5 \cdot 2^{5/3}/\pi^{1/3} = 4r \approx 10.8$ cm

5.7.25 $h = \dfrac{750}{\pi}\left(\dfrac{2\pi^2}{750^2}\right)^{1/3}$, $r = \left(\dfrac{750^2}{2\pi^2}\right)^{1/6}$

5.7.26 $h/r = \sqrt{2}$

5.7.27 $1/2$

5.7.28 \$7000

6.1.1 10

6.1.2 $35/3$

6.1.3 x^2

6.1.4 $2x^2$

6.1.5 $2x^2 - 8$

6.1.6 $2b^2 - 2a^2$

6.1.7 4 rectangles: $41/4 = 10.25$, 8 rectangles: $183/16 = 11.4375$

6.1.8 $23/4$

6.2.1 $87/2$

6.2.2 2

6.2.3 $\ln(10)$

6.2.4 $e^5 - 1$

6.2.5 $3^4/4$

6.2.6 $2^6/6 - 1/6$

6.2.7 $x^2 - 3x$

6.2.8 $2x(x^4 - 3x^2)$

6.2.9 e^{x^2}

6.2.10 $2xe^{x^4}$

6.2.11 $\tan(x^2)$

6.2.12 $2x\tan(x^4) - 10\tan(100x^2)$

6.2.13 31, 14

6.2.14 5

6.2.15 (a) 2/3

(b) 24/5

6.3.1 $(16/3)x^{3/2} + C$

6.3.2 $t^3 + t + C$

6.3.3 $8\sqrt{x} + C$

6.3.4 $-2/z + C$

6.3.5 $7\ln s + C$

6.3.6 $(5x+1)^3/15 + C$

6.3.7 $(x-6)^3/3 + C$

6.3.8 $2x^{5/2}/5 + C$

6.3.9 $-4/\sqrt{x} + C$

6.3.10 $4t - t^2 + C$, $t < 2$; $t^2 - 4t + 8 + C$, $t \geq 2$

7.1.1 $-(1-t)^{10}/10 + C$

7.1.2 $x^5/5 + 2x^3/3 + x + C$

7.1.3 $(x^2+1)^{101}/202 + C$

7.1.4 $-3(1-5t)^{2/3}/10 + C$

7.1.5 $(\sin^4 x)/4 + C$

7.1.6 $-(100-x^2)^{3/2}/3 + C$

7.1.7 $-2\sqrt{1-x^3}/3 + C$

7.1.8 $\sin(\sin \pi t)/\pi + C$

7.1.9 $1/(2\cos^2 x) = (1/2)\sec^2 x + C$

7.1.10 $-\ln|\cos x| + C$

7.1.11 0

7.1.12 $\tan^2(x)/2 + C$

7.1.13 $1/4$

7.1.14 $-\cos(\tan x) + C$

7.1.15 $1/10$

7.1.16 $\sqrt{3}/4$

7.1.17 $(27/8)(x^2-7)^{8/9}$

7.1.18 $-(3^7+1)/14$

7.1.19 0

7.1.20 $f(x)^2/2$

7.2.1 $x/2-\sin(2x)/4+C$

7.2.2 $-\cos x+(\cos^3 x)/3+C$

7.2.3 $3x/8-(\sin 2x)/4+(\sin 4x)/32+C$

7.2.4 $(\cos^5 x)/5-(\cos^3 x)/3+C$

7.2.5 $\sin x-(\sin^3 x)/3+C$

7.2.6 $(\sin^3 x)/3-(\sin^5 x)/5+C$

7.2.7 $-2(\cos x)^{5/2}/5+C$

7.2.8 $\tan x-\cot x+C$

7.2.9 $(\sec^3 x)/3-\sec x+C$

7.2.10 $-\cos x+\sin x+C$

7.2.11 $\dfrac{3}{2}\ln|\sec x+\tan x|+\tan x+\dfrac{1}{2}\sec x\tan x+C$

7.2.12 $\dfrac{\tan^5(x^2)}{10}+C$

7.3.1 $x\sqrt{x^2-1}/2-\ln|x+\sqrt{x^2-1}|/2+C$

7.3.2 $x\sqrt{9+4x^2}/2+(9/4)\ln|2x+\sqrt{9+4x^2}|+C$

7.3.3 $-(1-x^2)^{3/2}/3+C$

7.3.4 $\arcsin(x)/8-\sin(4\arcsin x)/32+C$

7.3.5 $\ln|x+\sqrt{1+x^2}|+C$

7.3.6 $(x+1)\sqrt{x^2+2x}/2-\ln|x+1+\sqrt{x^2+2x}|/2+C$

7.3.7 $-\arctan x-1/x+C$

7.3.8 $2\arcsin(x/2) - x\sqrt{4-x^2}/2 + C$

7.3.9 $\arcsin(\sqrt{x}) - \sqrt{x}\sqrt{1-x} + C$

7.3.10 $(2x^2+1)\sqrt{4x^2-1}/24 + C$

7.4.1 $\cos x + x\sin x + C$

7.4.2 $x^2\sin x - 2\sin x + 2x\cos x + C$

7.4.3 $(x-1)e^x + C$

7.4.4 $(1/2)e^{x^2} + C$

7.4.5 $(x/2) - \sin(2x)/4 + C =$
$(x/2) - (\sin x\cos x)/2 + C$

7.4.6 $x\ln x - x + C$

7.4.7 $(x^2\arctan x + \arctan x - x)/2 + C$

7.4.8 $-x^3\cos x + 3x^2\sin x + 6x\cos x - 6\sin x + C$

7.4.9 $x^3\sin x + 3x^2\cos x - 6x\sin x - 6\cos x + C$

7.4.10 $x^2/4 - (\cos^2 x)/4 - (x\sin x\cos x)/2 + C$

7.4.11 $x/4 - (x\cos^2 x)/2 + (\cos x\sin x)/4 + C$

7.4.12 $x\arctan(\sqrt{x}) + \arctan(\sqrt{x}) - \sqrt{x} + C$

7.4.13 $2\sin(\sqrt{x}) - 2\sqrt{x}\cos(\sqrt{x}) + C$

7.4.14 $\sec x\csc x - 2\cot x + C$

7.5.1 $-\ln|x-2|/4 + \ln|x+2|/4 + C$

7.5.2 $-x^3/3 - 4x - 4\ln|x-2| +$
$4\ln|x+2| + C$

7.5.3 $-1/(x+5) + C$

7.5.4 $-x - \ln|x-2| + \ln|x+2| + C$

7.5.5 $-4x + x^3/3 + 8\arctan(x/2) + C$

7.5.6 $(1/2)\arctan(x/2+5/2)+C$

7.5.7 $x^2/2-2\ln(4+x^2)+C$

7.5.8 $(1/4)\ln|x+3|-(1/4)\ln|x+7|+C$

7.5.9 $(1/5)\ln|2x-3|-(1/5)\ln|1+x|+C$

7.5.10 $(1/3)\ln|x|-(1/3)\ln|x+3|+C$

7.6.1 T,S: 4 ± 0

7.6.2 T: 9.28125 ± 0.281125; S: 9 ± 0

7.6.3 T: 60.75 ± 1; S: 60 ± 0

7.6.4 T: 1.1167 ± 0.0833; S: 1.1000 ± 0.0167

7.6.5 T: 0.3235 ± 0.0026; S: 0.3217 ± 0.000065

7.6.6 T: 0.6478 ± 0.0052; S: 0.6438 ± 0.000033

7.6.7 T: 2.8833 ± 0.0834; S: 2.9000 ± 0.0167

7.6.8 T: 1.1170 ± 0.0077; S: 1.1114 ± 0.0002

7.6.9 T: 1.097 ± 0.0147; S: 1.089 ± 0.0003

7.6.10 T: 3.63 ± 0.087; S: 3.62 ± 0.032

7.7.1 Converges to 1.

7.7.2 Diverges.

7.7.3 1/3

7.7.4 Divergent.

7.7.7 (a) $\pi/2$

(b) divergent (to ∞)

(c) 1

(d) divergent (to ∞)

(e) $\frac{5}{3}(4^{3/5})$

7.7.9 $0 < p < 1$

7.8.1 $\dfrac{(t+4)^4}{4}+C$

7.8.2 $\dfrac{(t^2-9)^{5/2}}{5}+C$

7.8.3 $\dfrac{(e^{t^2}+16)^2}{4}+C$

7.8.4 $\cos t - \dfrac{2}{3}\cos^3 t + C$

7.8.5 $\dfrac{\tan^2 t}{2}+C$

7.8.6 $\ln|t^2+t+3|+C$

7.8.7 $\dfrac{1}{8}\ln|1-4/t^2|+C$

7.8.8 $\dfrac{1}{25}\tan(\arcsin(t/5))+C = \dfrac{t}{25\sqrt{25-t^2}}+C$

7.8.9 $\dfrac{2}{3}\sqrt{\sin 3t}+C$

7.8.10 $t\tan t + \ln|\cos t| + C$

7.8.11 $2\sqrt{e^t+1}+C$

7.8.12 $\dfrac{3t}{8}+\dfrac{\sin 2t}{4}+\dfrac{\sin 4t}{32}+C$

7.8.13 $\dfrac{\ln|t|}{3}-\dfrac{\ln|t+3|}{3}+C$

7.8.14 $\dfrac{-1}{\sin\arctan t}+C = -\sqrt{1+t^2}/t+C$

7.8.15 $\dfrac{-1}{2(1+\tan t)^2}+C$

7.8.16 $\dfrac{(t^2+1)^{5/2}}{5}-\dfrac{(t^2+1)^{3/2}}{3}+C$

7.8.17 $\dfrac{e^t \sin t - e^t \cos t}{2} + C$

7.8.18 $\dfrac{(t^{3/2} + 47)^4}{6} + C$

7.8.19 $\dfrac{2}{3(2-t^2)^{3/2}} - \dfrac{1}{(2-t^2)^{1/2}} + C$

7.8.20 $\dfrac{\ln|\sin(\arctan(2t/3))|}{9} + C = (\ln(4t^2) - \ln(9+4t^2))/18 + C$

7.8.21 $\dfrac{(\arctan(2t))^2}{4} + C$

7.8.22 $\dfrac{3\ln|t+3|}{4} + \dfrac{\ln|t-1|}{4} + C$

7.8.23 $\dfrac{\cos^7 t}{7} - \dfrac{\cos^5 t}{5} + C$

7.8.24 $\dfrac{-1}{t-3} + C$

7.8.25 $\dfrac{-1}{\ln t} + C$

7.8.26 $\dfrac{t^2(\ln t)^2}{2} - \dfrac{t^2 \ln t}{2} + \dfrac{t^2}{4} + C$

7.8.27 $(t^3 - 3t^2 + 6t - 6)e^t + C$

7.8.28 $\dfrac{5+\sqrt{5}}{10}\ln(2t+1-\sqrt{5}) + \dfrac{5-\sqrt{5}}{10}\ln(2t+1+\sqrt{5}) + C$

8.1.1 It rises until $t = 100/49$, then falls. The position of the object at time t is $s(t) = -4.9t^2 + 20t + k$. The net distance traveled is $-45/2$, that is, it ends up $45/2$ meters below where it started. The total distance traveled is $6205/98$ meters.

8.1.2 $\displaystyle\int_0^{2\pi} \sin t \, dt = 0$

8.1.3 net: 2π, total: $2\pi/3 + 4\sqrt{3}$

8.1.4 8

8.1.5 $17/3$

8.1.6 $A = 18, B = 44/3, C = 10/3$

8.2.1 $8\sqrt{2}/15$

8.2.2 $1/12$

8.2.3 $9/2$

8.2.4 $4/3$

8.2.5 $2/3 - 2/\pi$

8.2.6 $3/\pi - 3\sqrt{3}/(2\pi) - 1/8$

8.2.7 $1/3$

8.2.8 $10\sqrt{5}/3 - 6$

8.2.9 $500/3$

8.2.10 2

8.2.11 $1/5$

8.2.12 $1/6$

8.3.5 $8\pi/3$

8.3.6 $\pi/30$

8.3.7 $\pi(\pi/2 - 1)$

8.3.8 (a) $114\pi/5$
(b) $74\pi/5$
(c) 20π
(d) 4π

8.3.9 $16\pi, 24\pi$

8.3.11 $\pi h^2(3r - h)/3$

8.3.13 2π

8.4.1 $2/\pi$; $2/\pi$; 0

8.4.2 $4/3$

8.4.3 $1/A$

8.4.4 $\pi/4$

8.4.5 $-1/3$, 1

8.4.6 $-4\sqrt{1224}$ ft/s; $-8\sqrt{1224}$ ft/s

8.5.1 $\approx 5,305,028,516$ N-m

8.5.2 $\approx 4,457,854,041$ N-m

8.5.3 $367,500\pi$ N-m

8.5.4 $49000\pi + 196000/3$ N-m

8.5.5 2450π N-m

8.5.6 0.05 N-m

8.5.7 $6/5$ N-m

8.5.8 3920 N-m

8.5.9 23520 N-m

8.5.10 12740 N-m

8.6.1 $15/2$

8.6.2 5

8.6.3 $16/5$

8.6.5 $\bar{x} = 45/28$, $\bar{y} = 93/70$

8.6.6 $\bar{x} = 0$, $\bar{y} = 4/(3\pi)$

8.6.7 $\bar{x} = 1/2$, $\bar{y} = 2/5$

8.6.8 $\bar{x} = 0$, $\bar{y} = 8/5$

8.6.9 $\bar{x} = 4/7$, $\bar{y} = 2/5$

8.6.10 $\bar{x} = \bar{y} = 1/5$

8.6.11 $\bar{x} = 0, \bar{y} = 28/(9\pi)$

8.6.12 $\bar{x} = \bar{y} = 28/(9\pi)$

8.6.13 $\bar{x} = 0, \bar{y} = 244/(27\pi) \approx 2.88$

8.7.1 $(22\sqrt{22} - 8)/27$

8.7.2 $\ln(2) + 3/8$

8.7.3 $a + a^3/3$

8.7.4 $\ln((\sqrt{2}+1)/\sqrt{3})$

8.7.6 $3/4$

8.7.7 ≈ 3.82

8.7.8 ≈ 1.01

8.7.9 $\sqrt{1+e^2} - \sqrt{2} + \frac{1}{2}\ln\left(\frac{\sqrt{1+e^2}-1}{\sqrt{1+e^2}+1}\right) + \frac{1}{2}\ln(3+2\sqrt{2})$

8.8.1 $8\pi\sqrt{3} - \frac{16\pi\sqrt{2}}{3}$

8.8.3 $\frac{730\pi\sqrt{730}}{27} - \frac{10\pi\sqrt{10}}{27}$

8.8.4 $\pi + 2\pi e + \frac{1}{4}\pi e^2 - \frac{\pi}{4e^2} - \frac{2\pi}{e}$

8.8.6 $8\pi^2$

8.8.7 $2\pi + \frac{8\pi^2}{3\sqrt{3}}$

8.8.8 $a > b$: $2\pi b^2 +$
$\frac{2\pi a^2 b}{\sqrt{a^2-b^2}} \arcsin(\sqrt{a^2-b^2}/a)$,
$a < b$: $2\pi b^2 +$
$\frac{2\pi a^2 b}{\sqrt{b^2-a^2}} \ln\left(\frac{b}{a} + \frac{\sqrt{b^2-a^2}}{a}\right)$

9.1.1 1

9.1.3 0

9.1.4 1

9.1.5 1

9.1.6 0

9.2.1 $\lim_{n\to\infty} n^2/(2n^2+1) = 1/2$

9.2.2 $\lim_{n\to\infty} 5/(2^{1/n}+14) = 1/3$

9.2.3 $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges, so $\displaystyle\sum_{n=1}^{\infty} 3\frac{1}{n}$ diverges

9.2.4 $-3/2$

9.2.5 11

9.2.6 20

9.2.7 $3/4$

9.2.8 $3/2$

9.2.9 $3/10$

9.3.1 diverges

9.3.2 diverges

9.3.3 converges

9.3.4 converges

9.3.5 converges

9.3.6 converges

9.3.7 diverges

9.3.8 converges

9.3.9 $N = 5$

9.3.10 $N = 10$

9.3.11 $N = 1687$

9.3.12 any integer greater than e^{200}

9.4.1 converges

9.4.2 converges

9.4.3 diverges

9.4.4 converges

9.4.5 0.90

9.4.6 0.95

9.5.1 converges

9.5.2 converges

9.5.3 converges

9.5.4 diverges

9.5.5 diverges

9.5.6 diverges

9.5.7 converges

9.5.8 diverges

9.5.9 converges

9.5.10 diverges

9.6.1 converges absolutely

9.6.2 diverges

9.6.3 converges conditionally

9.6.4 converges absolutely

9.6.5 converges conditionally

9.6.6 converges absolutely

9.6.7 diverges

9.6.8 converges conditionally

9.7.5 (a) converges

(b) converges

(c) converges

(d) diverges

9.8.1 (a) $R = 1, I = (-1, 1)$

(b) $R = \infty, I = (-\infty, \infty)$

(c) $R = e, I = (2 - e, 2 + e)$

(d) $R = 0$, converges only when $x = 2$

(e) $R = 1, I = [-6, -4]$

9.8.2 $R = e$

9.9.1 the alternating harmonic series

9.9.2 $\sum_{n=0}^{\infty} (n+1)x^n$

9.9.3 $\sum_{n=0}^{\infty} (n+1)(n+2)x^n$

9.9.4 $\sum_{n=0}^{\infty} \frac{(n+1)(n+2)}{2} x^n$, $R = 1$

9.9.5 $C + \sum_{n=0}^{\infty} \frac{-1}{(n+1)(n+2)} x^{n+2}$

9.10.1 (a) $\sum_{n=0}^{\infty} (-1)^n x^{2n}/(2n)!$, $R = \infty$

(b) $\sum_{n=0}^{\infty} x^n/n!$, $R = \infty$

(c) $\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{(x-5)^n}{5^{n+1}}$, $R=5$

(d) $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{(x-1)^n}{n}$, $R=1$

(e) $\displaystyle\ln(2)+\sum_{n=1}^{\infty}(-1)^{n-1}\frac{(x-2)^n}{n2^n}$, $R=2$

(f) $\displaystyle\sum_{n=0}^{\infty}(-1)^n(n+1)(x-1)^n$, $R=1$

(g) $\displaystyle 1+\sum_{n=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{n!2^n}x^n=1+\sum_{n=1}^{\infty}\frac{(2n-1)!}{2^{2n-1}(n-1)!\,n!}x^n$, $R=1$

(h) $x+x^3/3$

(i) $\displaystyle\sum_{n=0}^{\infty}(-1)^n x^{4n+1}/(2n)!$

(j) $\displaystyle\sum_{n=0}^{\infty}(-1)^n x^{n+1}/n!$

9.11.1 $\displaystyle 1-\frac{x^2}{2}+\frac{x^4}{24}-\frac{x^6}{720}+\cdots+\frac{x^{12}}{12!}$

9.11.2 1000; 8

9.11.3 $\displaystyle x+\frac{x^3}{3}+\frac{2x^5}{15}$, error ± 1.27.

10.1.2 $y=\arctan t+C$

10.1.3 $\displaystyle y=\frac{t^{n+1}}{n+1}+1$

10.1.4 $y=t\ln t-t+C$

10.1.5 $y=n\pi$, for any integer n.

10.1.6 none

10.1.7 $y=\pm\sqrt{t^2+C}$

10.1.8 $y=\pm 1$, $y=(1+Ae^{2t})/(1-Ae^{2t})$

10.1.9 $y^4/4-5y=t^2/2+C$

10.1.10 $y = (2t/3)^{3/2}$

10.1.11 $y = M + Ae^{-kt}$

10.1.12 $\dfrac{10\ln(15/2)}{\ln 5} \approx 2.52$ minutes

10.1.13 $y = \dfrac{M}{1 + Ae^{-Mkt}}$

10.1.14 $y = 2e^{3t/2}$

10.1.15 $t = -\dfrac{\ln 2}{k}$

10.1.16 $600e^{-6\ln 2/5} \approx 261$ mg; $\dfrac{5\ln 300}{\ln 2} \approx 41$ days

10.1.17 $100e^{-200\ln 2/191} \approx 48$ mg; $\dfrac{5730\ln 50}{\ln 2} \approx 32339$ years

10.1.18 $y = y_0 e^{t\ln 2}$

10.1.19 $500e^{-5\ln 2/4} \approx 210$ g

10.2.1 $y = Ae^{-5t}$

10.2.2 $y = Ae^{2t}$

10.2.3 $y = Ae^{-\arctan t}$

10.2.4 $y = Ae^{-t^3/3}$

10.2.5 $y = 4e^{-t}$

10.2.6 $y = -2e^{3t-3}$

10.2.7 $y = e^{1+\cos t}$

10.2.8 $y = e^2 e^{-e^t}$

10.2.9 $y = 0$

10.2.10 $y = 0$

10.2.11 $y = 4t^2$

10.2.12 $y = -2e^{(1/t)-1}$

10.2.13 $y = e^{1-t^{-2}}$

10.2.14 $y = 0$

10.2.15 $k = \ln 5$, $y = 100e^{-t\ln 5}$

10.2.16 $k = -12/13$, $y = \exp(-13t^{1/13})$

10.2.17 $y = 10^6 e^{t\ln(3/2)}$

10.2.18 $y = 10e^{-t\ln(2)/6}$

10.3.1 $y = Ae^{-4t} + 2$

10.3.2 $y = Ae^{2t} - 3$

10.3.3 $y = Ae^{-(1/2)t^2} + 5$

10.3.4 $y = Ae^{-e^t} - 2$

10.3.5 $y = Ae^t - t^2 - 2t - 2$

10.3.6 $y = Ae^{-t/2} + t - 2$

10.3.7 $y = At^2 - \dfrac{1}{3t}$

10.3.8 $y = \dfrac{c}{t} + \dfrac{2}{3}\sqrt{t}$

10.3.9 $y = A\cos t + \sin t$

10.3.10 $y = \dfrac{A}{\sec t + \tan t} + 1 - \dfrac{t}{\sec t + \tan t}$

10.4.1 $y(1) \approx 1.355$

10.4.2 $y(1) \approx 40.31$

10.4.3 $y(1) \approx 1.05$

10.4.4 $y(1) \approx 2.30$

10.5.1 $\dfrac{\omega+1}{2\omega}e^{\omega t} + \dfrac{\omega-1}{2\omega}e^{-\omega t}$

10.5.2 $2\cos(3t)+5\sin(3t)$

10.5.3 $-(1/4)e^{-5t}+(5/4)e^{-t}$

10.5.4 $-2e^{-3t}+2e^{4t}$

10.5.5 $5e^{-6t}+20te^{-6t}$

10.5.6 $(16t-3)e^{4t}$

10.5.7 $-2\cos(\sqrt{5}t)+\sqrt{5}\sin(\sqrt{5}t)$

10.5.8 $-\sqrt{2}\cos t+\sqrt{2}\sin t$

10.5.9 $e^{-6t}(4\cos t+24\sin t)$

10.5.10 $2e^{-3t}\sin(3t)$

10.5.11 $2\cos(2t-\pi/6)$

10.5.12 $5\sqrt{2}\cos(10t-\pi/4)$

10.5.13 $\sqrt{2}e^{-2t}\cos(3t-\pi/4)$

10.5.14 $5e^{4t}\cos(3t+\arcsin(4/5))$

10.5.15 $(2\cos(5t)+\sin(5t))e^{-2t}$

10.5.16 $-(1/2)e^{-2t}\sin(2t)$

10.6.1 $Ae^{5t}+Bte^{5t}+(6/169)\cos t-(5/338)\sin t$

10.6.2 $Ae^{-\sqrt{2}t}+Bte^{-\sqrt{2}t}+5$

10.6.3 $A\cos(4t)+B\sin(4t)+(1/2)t^2+(3/16)t-5/16$

10.6.4 $A\cos(\sqrt{2}t)+B\sin(\sqrt{2}t)-(\cos(5t)+\sin(5t))/23$

10.6.5 $e^t(A\cos t+B\sin t)+e^{2t}/2$

10.6.6 $Ae^{\sqrt{6}t}+Be^{-\sqrt{6}t}+2-t/3-e^{-t}/5$

10.6.7 $Ae^{-3t}+Be^{2t}-(1/5)te^{-3t}$

10.6.8 $Ae^t+Be^{3t}+(1/2)te^{3t}$

10.6.9 $A\cos(4t)+B\sin(4t)+(1/8)t\sin(4t)$

10.6.10 $A\cos(3t)+B\sin(3t)-(1/2)t\cos(3t)$

10.6.11 $Ae^{-6t}+Bte^{-6t}+3t^2e^{-6t}$

10.6.12 $Ae^{4t}+Bte^{4t}-t^2e^{4t}$

10.6.13 $Ae^{-t}+Be^{-5t}+(4/5)$

10.6.14 $Ae^{4t}+Be^{-3t}+(1/144)-(t/12)$

10.6.15 $A\cos(\sqrt{5}t)+B\sin(\sqrt{5}t)+8\sin(2t)$

10.6.16 $Ae^{2t}+Be^{-2t}+te^{2t}$

10.6.17 $4e^t+e^{-t}-3t-5$

10.6.18 $-(4/27)\sin(3t)+(4/9)t$

10.6.19 $e^{-6t}(2\cos t+20\sin t)+2e^{-4t}$

10.6.20 $\left(-\frac{23}{325}\cos(3t)+\frac{592}{975}\sin(3t)\right)+\frac{23}{325}\cos t-\frac{11}{325}\sin t$

10.6.21 $e^{-2t}(A\sin(5t)+B\cos(5t))+8\sin(2t)+25\cos(2t)$

10.6.22 $e^{-2t}(A\sin(2t)+B\cos(2t))+(14/195)\sin t-(8/195)\cos t$

10.7.1 $A\sin(t)+B\cos(t)-$
$\cos t\ln|\sec t+\tan t|$

10.7.2 $A\sin(t)+B\cos(t)+\frac{1}{5}e^{2t}$

10.7.3 $A\sin(2t)+B\cos(2t)+\cos t-\sin t\cos t\ln|\sec t+\tan t|$

10.7.4 $A\sin(2t)+B\cos(2t)+\frac{1}{2}\sin(2t)\sin^2(t)+\frac{1}{2}\sin(2t)\ln|\cos t|-\frac{t}{2}\cos(2t)+\frac{1}{4}\sin(2t)\cos(2t)$

10.7.5 $Ae^{2t}+Be^{-3t}+\frac{t^3}{15}e^{2t}-\left(\frac{t^2}{5}-\frac{2t}{25}+\frac{2}{125}\right)\frac{e^{2t}}{5}$

10.7.6 $Ae^t\sin t+Be^t\cos t-e^t\cos t\ln|\sec t+\tan t|$

10.7.7 $Ae^t\sin t+Be^t\cos t-\frac{1}{10}\cos t(\cos^3 t+3\sin^3 t-2\cos t-\sin t)+\frac{1}{10}\sin t(\sin^3 t-3\cos^3 t-2\sin t+$
$\cos t)=\frac{1}{10}\cos(2t)-\frac{1}{20}\sin(2t)$

11.1.2

(a) $\theta = \arctan(3)$
(b) $r = -4\csc\theta$
(c) $r = \sec\theta\csc^2\theta$
(d) $r = \sqrt{5}$
(e) $r^2 = \sin\theta\sec^3\theta$
(f) $r\sin\theta = \sin(r\cos\theta)$
(g) $r = 2/(\sin\theta - 5\cos\theta)$
(h) $r = 2\sec\theta$
(i) $0 = r^2\cos^2\theta - r\sin\theta + 1$

11.1.4 (a) $(x^2+y^2)^2 = 4x^2y - (x^2+y^2)y$
(b) $(x^2+y^2)^{3/2} = y^2$
(c) $x^2+y^2 = x^2y^2$
(d) $x^4+x^2y^2 = y^2$

11.2.1 (a) $(\theta\cos\theta+\sin\theta)/(-\theta\sin\theta+\cos\theta)$, $(\theta^2+2)/(-\theta\sin\theta+\cos\theta)^3$

(b) $\dfrac{\cos\theta+2\sin\theta\cos\theta}{\cos^2\theta-\sin^2\theta-\sin\theta}, \dfrac{3(1+\sin\theta)}{(\cos^2\theta-\sin^2\theta-\sin\theta)^3}$

(c) $(\sin^2\theta-\cos^2\theta)/(2\sin\theta\cos\theta)$, $-1/(4\sin^3\theta\cos^3\theta)$

(d) $\dfrac{2\sin\theta\cos\theta}{\cos^2\theta-\sin^2\theta}, \dfrac{2}{(\cos^2\theta-\sin^2\theta)^3}$

(e) undefined

(f) $\dfrac{2\sin\theta-3\sin^3\theta}{3\cos^3\theta-2\cos\theta}, \dfrac{3\cos^4\theta-3\cos^2\theta+2}{2\cos^3\theta(3\cos^2\theta-2)^3}$

11.3.1 (a) 1
(b) $9\pi/2$
(c) $\sqrt{3}/3$
(d) $\pi/12+\sqrt{3}/16$
(e) $\pi a^2/4$
(f) $41\pi/2$

11.3.2 $2-\pi/2$

11.3.3 $\pi/12$

11.3.4 $3\pi/16$

11.3.5 $\pi/4-3\sqrt{3}/8$

11.3.6 $\pi/2+3\sqrt{3}/8$

11.3.7 1

11.3.8 $3/2-\pi/4$

11.3.9 $\pi/3+\sqrt{3}/2$

11.3.10 $\pi/3-\sqrt{3}/4$

11.3.11 $4\pi^3/3$

11.3.12 π^2

11.3.13 $5\pi/24-\sqrt{3}/4$

11.3.14 $7\pi/12-\sqrt{3}$

11.3.15 $4\pi-\sqrt{15}/2-7\arccos(1/4)$

11.3.16 $3\pi^3$

11.4.6 $x=t-\dfrac{\sin(t)}{2},\ t=1-\dfrac{\cos(t)}{2}$

11.5.1 There is a horizontal tangent at all multiples of π.

11.5.2 $9\pi/4$

11.5.3 $\displaystyle\int_0^{2\pi}\frac{1}{2}\sqrt{5-4\cos t}\,dt$

12.1.6 3, $\sqrt{26}$, $\sqrt{29}$

12.1.7 $\sqrt{14}$, $2\sqrt{14}$, $3\sqrt{14}$.

12.1.8 $(x-1)^2+(y-1)^2+(z-1)^2=4$.

12.1.9 $(x-2)^2+(y+1)^2+(z-3)^2=25$.

12.1.11 $(x-2)^2+(y-1)^2+(z+1)^2=16$, $(y-1)^2+(z+1)^2=12$

12.2.6 $\sqrt{10}$, $\langle 0,-2\rangle$, $\langle 2,8\rangle$ 2, $2\sqrt{17}$, $\langle -2,-6\rangle$

12.2.7 $\sqrt{14}$, $\langle 0,4,0\rangle$, $\langle 2,0,6\rangle$ 4, $2\sqrt{10}$, $\langle -2,-4,-6\rangle$

12.2.8 $\sqrt{2}$, $\langle 0,-2,3\rangle$, $\langle 2,2,-1\rangle$ $\sqrt{13}$, 3, $\langle -2,0,-2\rangle$

12.2.9 $\sqrt{3}$, $\langle 1,-1,4\rangle$, $\langle 1,-1,-2\rangle$ $3\sqrt{2}$, $\sqrt{6}$, $\langle -2,2,-2\rangle$

12.2.10 $\sqrt{14}$, $\langle 2,1,0\rangle$, $\langle 4,3,2\rangle$ $\sqrt{5}$, $\sqrt{29}$, $\langle -6,-4,-2\rangle$

12.2.11 $\langle -3,-3,-11\rangle$, $\langle -3/\sqrt{139},-3/\sqrt{139},-11/\sqrt{139}\rangle$ $\langle -12/\sqrt{139},-12/\sqrt{139},-44/\sqrt{139}\rangle$

12.2.12 $\langle 0,0,0\rangle$

12.2.13 $\mathbf{0}$; $\langle -r\sqrt{3}/2, r/2\rangle$; $\langle 0,-12r\rangle$; where r is the radius of the clock

12.3.1 3

12.3.2 0

12.3.3 2

12.3.4 -6

12.3.5 42

12.3.6 $\sqrt{6}/\sqrt{7}$, ≈ 0.39

12.3.7 $-11\sqrt{14}\sqrt{29}/406$, ≈ 2.15

12.3.8 0, $\pi/2$

12.3.9 $1/2$, $\pi/3$

12.3.10 $-1/\sqrt{3}$, ≈ 2.19

12.3.11 $\arccos(1/\sqrt{3}) \approx 0.96$

12.3.12 $\sqrt{5}$, $\langle 1,2,0\rangle$.

12.3.13 $3\sqrt{14}/7$, $\langle 9/7, 6/7, 3/7\rangle$.

12.3.14 $\langle 0,5\rangle$, $\langle 5\sqrt{3},0\rangle$

12.3.15 $\langle 0, 15\sqrt{2}/2\rangle$,$\langle 15\sqrt{2}/2, 0\rangle$

12.3.16 Any vector of the form $\langle a, -7a/2, -2a\rangle$

12.3.17 $\langle 1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3}\rangle$

12.3.18 No.

12.3.19 Yes.

12.4.1 $\langle 1,-2,1\rangle$

12.4.2 $\langle 4,-6,-2\rangle$

12.4.3 $\langle -7,13,-9\rangle$

12.4.4 $\langle 0, -1, 0\rangle$

12.4.5 3

12.4.6 $21\sqrt{2}/2$

12.4.7 1

12.5.1 $(x-6)+(y-2)+(z-1)=0$

12.5.2 $4(x+1)+5(y-2)-(z+3)=0$

12.5.3 $(x-1)-(y-2)=0$

12.5.4 $-2(x-1)+3y-2z=0$

12.5.5 $4(x-1)-6y=0$

12.5.6 $x+3y=0$

12.5.7 $\langle 1,0,3\rangle + t\langle 0,2,1\rangle$

12.5.8 $\langle 1,0,3\rangle + t\langle 1,2,-1\rangle$

12.5.9 $t\langle 1,1,-1\rangle$

12.5.10 $-2/5$, $13/5$

12.5.12 neither

12.5.13 parallel

12.5.14 intersect at $(3,6,5)$

12.5.15 same line

12.5.19 $7/\sqrt{3}$

12.5.20 $4/\sqrt{14}$

12.5.21 $\sqrt{131}/\sqrt{14}$

12.5.22 $\sqrt{68}/3$

12.5.23 $\sqrt{42}/7$

12.5.24 $\sqrt{21}/6$

12.6.1 (a) $(\sqrt{2}, \pi/4, 1)$, $(\sqrt{3}, \pi/4, \arccos(1/\sqrt{3}))$

(b) $(7\sqrt{2}, 7\pi/4, 5)$, $(\sqrt{123}, 7\pi/4, \arccos(5/\sqrt{123})$

(c) $(1,1,1)$, $(\sqrt{2}, 1, \pi/4)$

(d) $(0,0,-\pi)$, $(\pi, 0, \pi)$

12.6.2 $r^2 + z^2 = 4$

12.6.3 $r\cos\theta = 0$

12.6.4 $r^2 + 2z^2 + 2z - 5 = 0$

12.6.5 $z = e^{-r^2}$

12.6.6 $z = r$

12.6.7 $\sin\theta = 0$

12.6.8 $1 = \rho\cos\phi$

12.6.9 $\rho = 2\sin\theta\sin\phi$.

12.6.10 $\rho\sin\phi = 2$

12.6.11 $\cos\phi = 1/\sqrt{2}$

12.6.13 $z = mr$; $\cot\phi = m$ if $m \neq 0$, $\phi = 0$ if $m = 0$

12.6.14 A sphere with radius $1/2$, center at $(0, 1/2, 0)$

12.6.15 $0 < \theta < \pi/2$, $0 < \phi < \pi/2$, $\rho > 0$; $0 < \theta < \pi/2$, $r > 0$, $z > 0$

13.1.1 $z = y^2$, $z = x^2$, $z = 0$, lines of slope 1

13.1.2 $z = |y|$, $z = |x|$, $z = 2|x|$, diamonds

13.1.3 $z = e^{-y^2}\sin(y^2)$, $z = e^{-x^2}\sin(x^2)$, $z = e^{-2x^2}\sin(2x^2)$, circles

13.1.4 $z = -\sin(y)$, $z = \sin(x)$, $z = 0$, lines of slope 1

13.1.5 $z = y^4$, $z = x^4$, $z = 0$, hyperbolas

13.1.6 (a) $\{(x,y) \mid |x| \leq 3 \text{ and } |y| \geq 2\}$

(b) $\{(x,y) \mid 1 \leq x^2+y^2 \leq 3\}$

(c) $\{(x,y) \mid x^2+4y^2 \leq 16\}$

13.2.1 No limit; use $x=0$ and $y=0$.

13.2.2 No limit; use $x=0$ and $x=y$.

13.2.3 No limit; use $x=0$ and $x=y$.

13.2.4 Limit is zero.

13.2.5 Limit is 1.

13.2.6 Limit is zero.

13.2.7 Limit is -1.

13.2.8 Limit is zero.

13.2.9 No limit; use $x=0$ and $y=0$.

13.2.10 Limit is zero.

13.2.11 Limit is -1.

13.2.12 Limit is zero.

13.3.1 $-2xy\sin(x^2y)$, $-x^2\sin(x^2y)+3y^2$

13.3.2 $(y^2-x^2y)/(x^2+y)^2$, $x^3/(x^2+y)^2$

13.3.3 $2xe^{x^2+y^2}$, $2ye^{x^2+y^2}$

13.3.4 $y\ln(xy)+y$, $x\ln(xy)+x$

13.3.5 $-x/\sqrt{1-x^2-y^2}$, $-y/\sqrt{1-x^2-y^2}$

13.3.6 $\tan y$, $x\sec^2 y$

13.3.7 $-1/(x^2y)$, $-1/(xy^2)$

13.3.8 $z=-2(x-1)-3(y-1)-1$

13.3.9 $z=1$

13.3.10 $z=6(x-3)+3(y-1)+10$

13.3.11 $z=(x-2)+4(y-1/2)$

13.3.12 $\mathbf{r}(t)=\langle 2,1,4\rangle+t\langle 2,4,-1\rangle$

13.3.16 height

13.4.1 $4xt\cos(x^2+y^2)+6yt^2\cos(x^2+y^2)$

13.4.2 $2xy\cos t+2x^2t$

13.4.3 $2xyt\cos(st)+2x^2s$, $2xys\cos(st)+2x^2t$

13.4.4 $2xy^2t-4yx^2s$, $2xy^2s+4yx^2t$

13.4.5 x/z, $3y/(2z)$

13.4.6 $-2x/z$, $-y/z$

13.4.7 (a) $V'=(nR-0.2V)/P$

(b) $P'=(nR+0.6P)/2V$

(c) $T'=(3P-0.4V)/(nR)$

13.5.1 $9\sqrt{5}/5$

13.5.2 $\sqrt{2}\cos 3$

13.5.3 $e\sqrt{2}(\sqrt{3}-1)/4$

13.5.4 $\sqrt{3}+5$

13.5.5 $-\sqrt{6}(2+\sqrt{3})/72$

13.5.6 $-1/5, 0$

13.5.7 $4(x-2)+8(y-1)=0$

13.5.8 $2(x-3)+3(y-2)=0$

13.5.9 $\langle -1,-1-\cos 1,-\cos 1\rangle$, $-\sqrt{2+2\cos 1+2\cos^2 1}$

13.5.10 Any direction perpendicular to $\nabla T = \langle 1,1,1\rangle$, for example, $\langle -1,1,0\rangle$

13.5.11 $2(x-1)-6(y-1)+6(z-3)=0$

13.5.12 $6(x-1)+3(y-2)+2(z-3)=0$

13.5.13 $\langle 2+4t, -3-12t, -1-8t\rangle$

13.5.14 $\langle 4+8t, 2+4t, -2-36t\rangle$

13.5.15 $\langle 4+8t, 2+20t, 6-12t\rangle$

13.5.16 $\langle 0,1\rangle$, $\langle 4/5, -3/5\rangle$

13.5.18 (a) $\langle 4,9\rangle$

(b) $\langle -81,2\rangle$ or $\langle 81,-2\rangle$

13.5.19 in the direction of $\langle 8,1\rangle$

13.5.20 $\nabla g(-1,3) = \langle 2,1\rangle$

13.6.1 $f_{xx} = (2x^3y-6xy^3)/(x^2+y^2)^3$, $f_{yy} = (2xy^3-6x^3y)/(x^2+y^2)^3$

13.6.2 $f_x = 3x^2y^2$, $f_y = 2x^3y+5y^4$, $f_{xx} = 6xy^2$, $f_{yy} = 2x^3+20y^3$, $f_{xy} = 6x^2y$

13.6.3 $f_x = 12x^2+y^2$, $f_y = 2xy$,
$f_{xx} = 24x$, $f_{yy} = 2x$, $f_{xy} = 2y$

13.6.4 $f_x = \sin y$, $f_y = x\cos y$, $f_{xx} = 0$, $f_{yy} = -x\sin y$, $f_{xy} = \cos y$

13.6.5 $f_x = 3\cos(3x)\cos(2y)$,
$f_y = -2\sin(3x)\sin(2y)$,
$f_{xy} = -6\cos(3x)\sin(2y)$,
$f_{yy} = -4\sin(3x)\cos(2y)$,
$f_{xx} = -9\sin(3x)\cos(2y)$

13.6.6 $f_x = e^{x+y^2}$, $f_y = 2ye^{x+y^2}$,
$f_{xx} = e^{x+y^2}$,
$f_{yy} = 4y^2e^{x+y^2}+2e^{x+y^2}$,
$f_{xy} = 2ye^{x+y^2}$

13.6.7 $f_x = \dfrac{3x^2}{2(x^3+y^4)}$, $f_y = \dfrac{2y^3}{x^3+y^4}$, $f_{xx} = \dfrac{3x}{x^3+y^4} - \dfrac{9x^4}{2(x^3+y^4)^2}$, $f_{yy} = \dfrac{6y^2}{x^3+y^4} - \dfrac{8y^6}{(x^3+y^4)^2}$,
$f_{xy} = \dfrac{-6x^2y^3}{(x^3+y^4)^2}$

13.6.8 $z_x = \dfrac{-x}{16z}$, $z_y = \dfrac{-y}{4z}$,
$z_{xx} = -\dfrac{16z^2+x^2}{16^2z^3}$,
$z_{yy} = -\dfrac{4z^2+y^2}{16z^3}$,
$z_{xy} = \dfrac{-xy}{64z^3}$

13.6.9 $z_x = -\dfrac{y+z}{x+y}$, $z_y = -\dfrac{x+z}{x+y}$,
$z_{xx} = 2\dfrac{y+z}{(x+y)^2}$, $z_{yy} = 2\dfrac{x+z}{(x+y)^2}$,
$z_{xy} = \dfrac{2z}{(x+y)^2}$

13.7.1 minimum at $(1,-1)$

13.7.2 none

13.7.3 none

13.7.4 maximum at $(1,-1/6)$

13.7.5 none

13.7.6 minimum at $(2,-1)$

13.7.7 $f(2,2)=-2$, $f(2,0)=4$

13.7.8 a cube $1/\sqrt[3]{2}$ on a side

13.7.9 $65/3\times 65/3\times 130/3$

13.7.10 It has a square base, and is one and one half times as tall as wide. If the volume is V the dimensions are $\sqrt[3]{2V/3}\times\sqrt[3]{2V/3}\times\sqrt[3]{9V/4}$.

13.7.11 $\sqrt{100/3}$

13.7.12 $|ax_0+by_0+cz_0-d|/\sqrt{a^2+b^2+c^2}$

13.7.13 The sides and bottom should all be $2/3$ meter, and the sides should be bent up at angle $\pi/3$.

13.7.14 $(3,4/3)$

13.7.16 $|b|$ if $b\le 1/2$, otherwise $\sqrt{b-1/4}$

13.7.17 $|b|$ if $b \leq 1/2$, otherwise $\sqrt{b-1/4}$

13.7.19 $1024/\sqrt{3}$

13.8.1 a cube, $\sqrt[3]{1/2} \times \sqrt[3]{1/2} \times \sqrt[3]{1/2}$

13.8.2 $65/3 \cdot 65/3 \cdot 130/3 = 2 \cdot 65^3/27$

13.8.3 It has a square base, and is one and one half times as tall as wide. If the volume is V the dimensions are $\sqrt[3]{2V/3} \times \sqrt[3]{2V/3} \times \sqrt[3]{9V/4}$.

13.8.4 $|ax_0 + by_0 + cz_0 - d|/\sqrt{a^2+b^2+c^2}$

13.8.5 $(0,0,1)$, $(0,0,-1)$

13.8.6 $\sqrt[3]{4V} \times \sqrt[3]{4V} \times \sqrt[3]{V/16}$

13.8.7 Farthest: $(-\sqrt{2}, \sqrt{2}, 2+2\sqrt{2})$; closest: $(2,0,0)$, $(0,-2,0)$

13.8.8 $x = y = z = 16$

13.8.9 $(1,2,2)$

13.8.10 $(\sqrt{5},0,0)$, $(-\sqrt{5},0,0)$

13.8.11 standard \$65, deluxe \$75

13.8.12 $x = 9$, $\phi = \pi/3$

13.8.13 35, -35

13.8.14 maximum e^4, no minimum

13.8.15 5, $-9/2$

13.8.16 3, 3, 3

13.8.17 a cube of side length $2/\sqrt{3}$

14.1.1 16

14.1.2 4

14.1.3 $15/8$

14.1.4 $1/2$

14.1.5 $5/6$

14.1.6 $12-65/(2e)$.

14.1.7 $1/2$

14.1.8 $\pi/64$

14.1.9 $(2/9)2^{3/2}-(2/9)$

14.1.10 $(1-\cos(1))/4$

14.1.11 $(2\sqrt{2}-1)/6$

14.1.12 $\pi-2$

14.1.13 $(e^9-1)/6$

14.1.14 $\frac{4}{15}-\frac{\pi}{4}$

14.1.15 $1/3$

14.1.16 448

14.1.17 $4/5$

14.1.18 8π

14.1.19 2

14.1.20 $5/3$

14.1.21 $81/2$

14.1.22 $2a^3/3$

14.1.23 4π

14.1.24 $\pi/32$

14.1.25 $31/8$

14.1.26 $128/15$

14.1.27 1800π m^3

14.1.28 $\dfrac{(e^2+8e+16)}{15}\sqrt{e+4}-\dfrac{5\sqrt{5}}{3}-\dfrac{e^{5/2}}{15}+\dfrac{1}{15}$

14.1.30 $16-8\sqrt{2}$

14.2.1 4π

14.2.2 $32\pi/3-4\sqrt{3}\pi$

14.2.3 $(2-\sqrt{2})\pi/3$

14.2.4 $4/9$

14.2.5 $5\pi/3$

14.2.6 $\pi/6$

14.2.7 $\pi/2$

14.2.8 $\pi/2-1$

14.2.9 $\sqrt{3}/4+\pi/6$

14.2.10 $8+\pi$

14.2.11 $\pi/12$

14.2.12 $(1-\cos(9))\pi/2$

14.2.13 $-a^5/15$

14.2.14 12π

14.2.15 π

14.2.16 $16/3$

14.2.17 21π

3 2π

14.3.1 $\bar{x}=\bar{y}=2/3$

14.3.2 $\bar{x}=4/5, \bar{y}=8/15$

14.3.3 $\bar{x} = 0, \bar{y} = 3\pi/16$

14.3.4 $\bar{x} = 0, \bar{y} = 16/(15\pi)$

14.3.5 $\bar{x} = 3/2, \bar{y} = 9/4$

14.3.6 $\bar{x} = 6/5, \bar{y} = 12/5$

14.3.7 $\bar{x} = 14/27, \bar{y} = 28/55$

14.3.8 $(3/4, 2/5)$

14.3.9 $\left(\dfrac{81\sqrt{3}}{80\pi}, 0\right)$

14.3.10 $\bar{x} = \pi/2, \bar{y} = \pi/8$

14.3.11 $M = \displaystyle\int_0^{2\pi}\int_0^{1+\cos\theta}(2+\cos\theta)r\,dr\,d\theta,$
$M_x = \displaystyle\int_0^{2\pi}\int_0^{1+\cos\theta}\sin\theta(2+\cos\theta)r^2\,dr\,d\theta,$
$M_y = \displaystyle\int_0^{2\pi}\int_0^{1+\cos\theta}\cos\theta(2+\cos\theta)r^2\,dr\,d\theta.$

14.3.12 $M = \displaystyle\int_{-\pi/2}^{\pi/2}\int_0^{\cos\theta}(r+1)r\,dr\,d\theta,$
$M_x = \displaystyle\int_{-\pi/2}^{\pi/2}\int_0^{\cos\theta}\sin\theta(r+1)r^2\,dr\,d\theta,$
$M_y = \displaystyle\int_{-\pi/2}^{\pi/2}\int_0^{\cos\theta}\cos\theta(r+1)r^2\,dr\,d\theta.$

14.3.13 $M = \displaystyle\int_{-\pi/2}^{\pi/2}\int_{\cos\theta}^{1+\cos\theta}r\,dr\,d\theta + \int_{\pi/2}^{3\pi/2}\int_0^{1+\cos\theta}r\,dr\,d\theta,$
$M_x = \displaystyle\int_{-\pi/2}^{\pi/2}\int_{\cos\theta}^{1+\cos\theta}r^2\sin\theta\,dr\,d\theta + \int_{\pi/2}^{3\pi/2}\int_0^{1+\cos\theta}r^2\sin\theta\,dr\,d\theta,$
$M_y = \displaystyle\int_{-\pi/2}^{\pi/2}\int_{\cos\theta}^{1+\cos\theta}r^2\cos\theta\,dr\,d\theta + \int_{\pi/2}^{3\pi/2}\int_0^{1+\cos\theta}r^2\cos\theta\,dr\,d\theta.$

14.4.1 $\pi a\sqrt{h^2+a^2}$

14.4.2 $\pi a^2\sqrt{m^2+1}$

14.4.3 $\sqrt{3}/2$

14.4.4 $\pi\sqrt{2}$

14.4.5 $\pi\sqrt{2}/8$

14.4.6 $\pi/2-1$

14.4.7 $\dfrac{d^2\sqrt{a^2+b^2+c^2}}{2abc}$

14.4.8 $8\sqrt{3}\pi/3$

14.5.1 $11/24$

14.5.2 $623/60$

14.5.3 $-3e^2/4+2e-3/4$

14.5.4 $1/20$

14.5.5 $\pi/48$

14.5.6 $11/84$

14.5.7 $151/60$

14.5.8 π

14.5.10 $\dfrac{3\pi}{16}$

14.5.11 32

14.5.12 $64/3$

14.5.13 $\bar{x}=\bar{y}=0$, $\bar{z}=16/15$

14.5.14 $\bar{x}=\bar{y}=0$, $\bar{z}=1/3$

14.6.1 $\pi/12$

14.6.2 $\pi(1-\sqrt{2}/2)$

14.6.3 $5\pi/4$

14.6.4 0

14.6.5 $5\pi/4$

14.6.6 $4/5$

14.6.7 $256\pi/15$

14.6.8 $4\pi^2$

14.6.9 $\dfrac{3\pi}{16}$

14.6.10 $\pi k h^2 a^2/12$

14.6.11 $\pi k h a^3/6$

14.6.12 $\pi^2/4$

14.6.13 $4\pi/5$

14.6.14 15π

14.6.15 $9k\pi(5\sqrt{2}-2\sqrt{5})/20$

14.7.1 $4\pi\sqrt{3}/3$

14.7.2 0

14.7.3 $2/3$

14.7.4 $\dfrac{e^2-1}{2e^2}$

14.7.5 36

14.7.6 $32(\sqrt{2}+\ln(1+\sqrt{2}))/3$

14.7.7 $3\cos(1)-3\cos(4)$

14.7.8 $\pi(1-\cos(1))/24$

14.7.10 $(4/3)\pi abc$

15.1.5 $\langle 3\cos t, 3\sin t, 2-3\sin t\rangle$

15.1.6 $\langle 0, t\cos t, t\sin t\rangle$

15.2.1 $\langle 2t, 0, 1\rangle$, $\mathbf{r}'/\sqrt{1+4t^2}$

15.2.2 $\langle -\sin t, 2\cos 2t, 2t\rangle$, $\mathbf{r}'/\sqrt{\sin^2 t + 4\cos^2(2t) + 4t^2}$

15.2.3 $\langle -e^t \sin(e^t), e^t \cos(e^t), \cos t\rangle$, $\mathbf{r}'/\sqrt{e^{2t} + \cos^2 t}$

15.2.4 $\langle \sqrt{2}/2, \sqrt{2}/2, \pi/4\rangle + t\langle -\sqrt{2}/2, \sqrt{2}/2, 1\rangle$

15.2.5 $\langle 1/2, \sqrt{3}/2, -1/2\rangle + t\langle -\sqrt{3}/2, 1/2, 2\sqrt{3}\rangle$

15.2.6 $2/\sqrt{5}/\sqrt{4+\pi^2}$

15.2.7 $7\sqrt{5}\sqrt{17}/85$, $-9\sqrt{5}\sqrt{17}/85$

15.2.9 $\langle 0, t\cos t, t\sin t\rangle$, $\langle 0, \cos t - t\sin t, \sin t + t\cos t\rangle$, $\mathbf{r}'/\sqrt{1+t^2}$, $\sqrt{1+t^2}$

15.2.10 $\langle \sin t, 1-\cos t, t^2/2\rangle$

15.2.11 $t = 4$

15.2.12 37, 1

15.2.13 $\langle t^2/2, t^3/3, \sin t\rangle$

15.2.16 $(1,1,1)$ when $t = 1$ and $s = 0$; $\theta = \arccos(3/\sqrt{14})$; no

15.2.17 $-6x + (y-\pi) = 0$

15.2.18 $-x/\sqrt{2} + y/\sqrt{2} + 6z = 0$

15.2.19 $(-1,-3,1)$

15.2.20 $\langle 1/\sqrt{2}, 1/\sqrt{2}, 0\rangle + t\langle -1, 1, 6\sqrt{2}\rangle$

15.3.1 $2\pi\sqrt{13}$

15.3.2 $(-8 + 13\sqrt{13})/27$

15.3.3 $\sqrt{5}/2 + \ln(\sqrt{5}+2)/4$

15.3.4 $(85\sqrt{85} - 13\sqrt{13})/27$

15.3.5 $\int_0^5 \sqrt{1+e^{2t}}\, dt$

15.4.1 $2\sqrt{2}/(2+4t^2)^{3/2}$

15.4.2 $2\sqrt{2}/(1+8t^2)^{3/2}$

15.4.3 $\sqrt{3600t^{10}+400t^6+36t^2}/(1+9t^4+25t^8)^{3/2}$

15.4.4 $12\sqrt{17}/289$

15.5.1 $\langle 5t^4, 4t, 1\rangle$, $\langle 20t^3, 4, 0\rangle$, $a_T = 100t^7+16t/\sqrt{25t^8+16t^2}$, $a_N = \sqrt{3600t^8+400t^6-16}/\sqrt{25t^8+16t^2}$

15.5.2 $\langle -\sin t, \cos t, 2t\rangle$, $\langle -\cos t, -\sin t, 2\rangle$, $4t/\sqrt{4t^2+1}$, $\sqrt{4t^2+5}/\sqrt{4t^2+1}$

15.5.3 $\langle -\sin t, \cos t, e^t\rangle$, $\langle -\cos t, -\sin t, e^t\rangle$, $e^{2t}/\sqrt{e^{2t}+1}$, $\sqrt{2e^{2t}+1}/\sqrt{e^{2t}+1}$

15.5.4 $\langle e^t, \cos t, e^t\rangle$, $\langle e^t, -\sin t, e^t\rangle$, $(2e^{2t}-\cos t\sin t)/\sqrt{2e^{2t}+\cos^2 t}$, $\sqrt{2}e^t|\cos t+\sin t|/\sqrt{2e^{2t}+\cos^2 t}$

15.5.5 $\langle -3\sin t, 2\cos t, 0\rangle$, $\langle 3\cos t, 2\sin t, 0\rangle$

15.5.6 $\langle -3\sin t, 2\cos t+0.1, 0\rangle$, $\langle 3\cos t, 2\sin t+t/10, 0\rangle$

15.5.7 $\langle -3\sin t, 2\cos t, 1\rangle$, $\langle 3\cos t, 2\sin t, t\rangle$

15.5.8 $\langle -3\sin t, 2\cos t+1/10, 1\rangle$, $\langle 3\cos t, 2\sin t+t/10, t\rangle$

16.2.1 -1, 0

16.2.2 0, $a+b$

16.2.3 $(2b-a)/3$, 0

16.2.4 0, 1

16.2.5 -2π, 0

16.2.6 0, 2π

16.3.1 $13\sqrt{11}/4$

16.3.2 0

16.3.3 $3\sin(4)/2$

16.3.4 $2e^3$

16.3.5 128

16.3.6 $(9e-3)/2$

16.3.7 $e^{e+1}-e^e-e^{1/e-1}+e^{1/e}+e^4/4-e^{-4}/4$

16.3.8 $1+\sin(1)-\cos(1)$

16.3.9 $3\ln 3-2\ln 2$

16.3.10 $3/20+10\ln(2)/7$

16.3.11 $2\ln 5-2\ln 2+15/32$

16.3.12 1

16.3.13 0

16.3.14 $21+\cos(1)-\cos(8)$

16.3.15 $(\ln 29-\ln 2)/2$

16.3.16 $2\ln 2+\pi/4-2$

16.3.17 $1243/3$

16.3.18 $\ln 2+11/3$

16.3.19 $3\cos(1)-\cos(2)-\cos(4)-\cos(8)$

16.3.20 $-10/3$

16.3.22 no f

16.3.23 $x^4/4-y^5/5$

16.3.24 no f

16.3.25 no f

16.3.26 $y\sin x$

16.3.27 no f

16.3.28 xyz

16.3.29 414

16.3.30 6

16.3.31 $1/e-\sin 3$

16.3.32 $1/\sqrt{77} - 1/\sqrt{3}$

16.4.1 1

16.4.2 0

16.4.3 $1/(2e) - 1/(2e^7) + e/2 - e^7/2$

16.4.4 $1/2$

16.4.5 $-1/6$

16.4.6 $(2\sqrt{3} - 10\sqrt{5} + 8\sqrt{6})/3 - 2\sqrt{2}/5 + 1/5$

16.4.7 $11/2 - \ln(2)$

16.4.8 $2 - \pi/2$

16.4.9 $-17/12$

16.4.10 0

16.4.11 $-\pi/2$

16.4.12 12π

16.4.13 $2\cos(1) - 2\sin(1) - 1$

16.5.1 both are $-45\pi/4$

16.5.2 $a^2bc + ab^2c + abc^2$

16.5.3 $e^2 - 2e + 7/2$

16.5.4 3

16.5.5 $384\pi/5$

16.5.6 $\pi/3$

16.5.7 10π

16.5.8 $\pi/2$

16.6.2 $25\sqrt{21}/4$

16.6.3 $\pi\sqrt{21}$

16.6.4 $\pi(5\sqrt{5}-1)/6$

16.6.5 $4\pi\sqrt{2}$

16.6.6 $\pi a^2/2$

16.6.7 $2\pi a(a-\sqrt{a^2-b^2})$

16.6.8 $\pi((1+4a^2)^{3/2}-1)/6$

16.6.9 $\pi a^2 - 2a^2$

16.6.10 $\pi a^2\sqrt{1+k^2}/4$

16.6.11 $A\sqrt{1+a^2+b^2}$

16.6.12 $A\sqrt{k^2+1}$

16.6.13 $8a^2$

16.7.1 $(0,0,3/8)$

16.7.2 $(11/20,11/20,3/10)$

16.7.3 $(0,0,1364/425)$

16.7.4 on center axis, $h/3$ above the base

16.7.5 16

16.7.6 7

16.7.7 $-\pi$

16.7.8 $-2/e$

16.7.9 $\pi b^2(-4b^4-3b^2+6a^2b^2+6a^2)/6$

16.7.10 9280 kg/s

16.7.11 $24\varepsilon_0$

16.8.1 -3π

16.8.2 0

16.8.3 -4π

16.8.4 3π

16.8.5 $A(p(c-b)+q(a-c)+a-b)$

Index