Calculus Notes

West Virginia Wesleyan College

Calculus Notes

West Virginia Wesleyan College

Jesse Oldroyd
West Virginia Wesleyan College

November 10, 2022

Go Seahawks.

Edition: Annual Edition 2022

Website: j-oldroyd.github.io/wvwc-calculus

Preface

This document was created to serve as a single source for my lecture notes for the calculus sequence at West Virginia Wesleyan College. As such these notes are divided into three self-explanatory parts.

- The first part, Differential Calculus, introduces the derivative and covers its important properties and applications. This content forms the majority of the Calculus I course.
- The next part, Integral Calculus, introduces the integral and goes over multiple methods for its calculation. This content forms the majority of the Calculus II course.
- The final part, Multivariable Calculus, generalizes the concepts of derivatives and integrals to two and three dimensions. This content forms basis of the Calculus III course.

This document is very much *in progress* and therefore typos and other errors are to be expected. If you find any, I would appreciate you letting me know by contacting me by email.

Contents

Part I

Differential Calculus

Chapter 1

Functions

The primary object of study in calculus is the function. In this chapter, we'll review important types of functions that will appear in this course.

1.1 Function Review

This section reviews basic facts about functions. To appear in a later version.

1.2 Types of Functions

Linear Functions

The most basic example of a function that we'll study in calculus is the *linear function*.

Definition 1.2.1 Linear Functions. A **linear function** is a function $f(x)$ of the form

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

for constants m and b. ◊

As the name suggests, any (non-vertical) line is an example of a linear function. Such functions are completely determined by the slope of the corresponding line and the y-intercept.

Equivalently, any line is determined by knowing two distinct points on the line.

In particular, if $y = mx + b$ then m is the slope and b is the y-intercept. This is known as the *slope-intercept* equation of a line. Equations of lines are also often written in *point-slope form* as $y - y_0 = m(x - x_0)$.

Example 1.2.2 Finding the Equation of a Line. Find a function $f(x)$ whose graph is a line passing through $(-5, 2)$ and $(1, 1)$. □

Polynomial Functions

After linear functions, we also study functions with higher powers of x.

Definition 1.2.3 Polynomial Functions. A **polynomial function** is a function of the form $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ where n is a nonnegative whole number and the **coefficients** $a_0, a_1, \ldots, a_n$ are constants and n is the **degree** of the polynomial. ◊

If $f(x) = 0$, then we say that $f(x)$ has degree $-\infty$.

So linear functions are just polynomial functions of degree at most 1. And just as linear functions are determined by knowing two distinct points on the line, any polynomial functions is determined by knowing $n+1$ distinct points on the polynomial.

Example 1.2.4 Composition of Polynomial Functions. Let $f(x) = 3x - x^2$ and $g(x) = 4x^3 + 5$. Find $f(g(x))$. □

Algebraic Functions

The operations of addition, subtraction, multiplication, division, taking whole number powers and taking whole number roots applied to polynomials give rise to *algebraic functions*. Two particularly important examples are *rational functions* and *root functions*.

Definition 1.2.5 Rational and Root Functions. A **rational function** is a function of the form $\frac{p(x)}{q(x)}$ where p and q are polynomial functions. A **root function** is a function of the form $x^{1/n} = \sqrt[n]{x}$ for some natural number n. ◊

As functions become more complicated, we have to worry more and more about their domains. For a polynomial function, the domain is the set of all real numbers $\mathbb{R}$ (if we ignore complex numbers). The domain of a rational function is the set of all numbers where the denominator is nonzero. The domain of a root function is the set of all nonnegative numbers.

Example 1.2.6 Finding the Domain of an Algebraic Function. Find the domain of

$$\sqrt{x^2 - 3x + 2} + \frac{x-4}{3x+4}.$$

Solution. We need to find where both the radicand $x^2 - 3x + 2$ is nonnegative and where $3x + 4 \neq 0$. If we solve $x^2 - 3x + 2 \geq 0$, we see that x must be in $(-\infty, 1] \cup [2, \infty)$. Likewise, $3x + 4 \neq 0$ if and only if $x \neq -\frac{4}{3}$. Hence the domain is

$$(-\infty, -\frac{4}{3}) \cup (-\frac{4}{3}, 1] \cup [2, \infty).$$

□

We'll make use of the computer algebra system Sage to perform certain computations. As an example, we use it below to solve the inequality $x^2 - 3x + 2 \geq 0$:

```
# click the Sage button to evaluate this cell
solve(x^2 - 3*x + 2 >= 0, x)    # tells Sage the inequality
    to solve and the variable to solve for
```

1.3 Trigonometric Functions

Angles and Terminal Points

Consider the unit circle in $\mathbb{R}^2$, which is given by $x^2 + y^2 = 1$. Each point P on this circle makes an angle θ with the positive x-axis, and therefore θ must also determine the point completely. The angle θ is typically specified using either *radians* or *degrees*. Converting from one unit to the other can be done by noting that π radians is precisely equal to 180 degrees.

Unless otherwise specified, we will use radians for angle measures in this course.

Example 1.3.1 Converting from Degrees to Radians. Convert 270 degrees to radians and $\frac{\pi}{6}$ radians to degrees. □

Just as points on the unit circle determine angles, angles also determine points. A point P determined by an angle θ is also known as the *terminal point* for the angle. Terminal points for certain *reference angles* are useful to remember.

Example 1.3.2 Finding Terminal Points. Find the terminal points for the angles $210°$ and $\frac{5\pi}{3}$.

Solution. First, note that $210° = \frac{7\pi}{6}$. We therefore choose $\frac{\pi}{6}$ as our reference angle and obtain the terminal point $(-\frac{\sqrt{3}}{2}, -\frac{1}{2})$. The terminal point for $\frac{5\pi}{3}$ is likewise $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$. □

The Trigonometric Functions

Since angles determine terminal points on circles, the coordinates (x, y) of each point on the circle can be viewed as functions of the angle θ. These *coordinate functions*, $x(\theta)$ and $y(\theta)$, are the fundamental trigonometric functions *sine* and *cosine*, and can be used to define the other four trigonometric functions commonly used.

Definition 1.3.3 Trigonometric Functions. Let $\theta \in \mathbb{R}$ and let $P = (x, y)$ denote the corresponding terminal point. The **cosine** function is the function $\cos\theta = x$, and the **sine** function is the function $\sin\theta = y$. The **secant, cosecant, tangent** and **cotangent** functions are defined as follows:

$$\sec\theta = \frac{1}{\cos\theta}, \csc\theta = \frac{1}{\sin\theta}, \tan\theta = \frac{\sin\theta}{\cos\theta} \text{ and } \cot\theta = \frac{\cos\theta}{\sin\theta}.$$

♢

The trigonometric functions satisfy important equalities known as *Pythagorean identities*.

Theorem 1.3.4 Pythagorean Identities. *Let $\theta \in \mathbb{R}$. Then*

$$\sin^2\theta + \cos^2\theta = 1.$$

If $\theta \neq \frac{\pi}{2} + k\pi$ for some $k \in \mathbb{Z}$, then

$$\tan^2\theta + 1 = \sec^2\theta.$$

If $\theta \neq k\pi$ for some $k \in \mathbb{Z}$, then

$$\cot^2\theta + 1 = \csc^2\theta.$$

Chapter 2

Limits of Functions

At its heart, calculus is just the mathematics of change. In particular, calculus provides the tools necessary to describe how a function changes. Since a function can be used to represent many quantities that appear in real life, calculus therefore gives us a way to study how many different quantities (such as temperature, acceleration, energy of a signal, etc.) can change. But before we can start using calculus, we need to come up with the proper language to describe it. The language we will need to develop is that of the *limit of a function*.

2.1 The Limit of a Function

Motivating Limits

Imagine creating a mathematical valley out of the graph of $f(x) = x^2$, and in this valley walks a mathematical ant. The ant is walking towards the place on the hill directly above $x = -\frac{1}{2}$. At $x = -2$, the ant is 4 units above the ground. At $x = -1$, the ant is now 1 unit above the ground. As the ant moves towards $x = -\frac{1}{2}$, its height above the ground gets closer and closer to $\frac{1}{4}$. To say it more clearly, and because modern calculus is built on this idea, *as the ant approaches* the point above $x = -\frac{1}{2}$, its *height above the ground approaches* the value $\frac{1}{4}$. In other words, the *limit* of $f(x) = x^2$ as x approaches $-\frac{1}{2}$ is $\frac{1}{4}$.

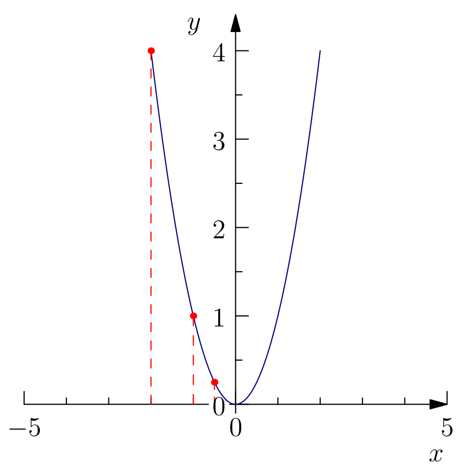


Figure 2.1.1 A mathematical valley minus a mathematical ant.

Remember that a function f of the variable x is just a rule that turns one number, x, into another number, $f(x)$. So the idea that the *limit of a function* is trying to express is what happens to the number $f(x)$ (the output) as the number x (the input) approaches some particular value.

Actually, functions are much more general than this. But for calculus, it won't hurt us to view functions in this way.

We're not quite ready to define the limit of a function precisely, but we can point one thing out right away: *the limit of a function requires two pieces of information: the function itself and the number that x is approaching.* The limit should then be whatever number that $f(x)$ is approaching.

Example 2.1.2 Estimating the limit of a trigonometric function. Let $f(x) = \sin x \cos x$. What is the limit of $f(x)$ as x approaches the number π?

Solution. We don't have a lot of tools to find limits yet, so we'll try to estimate it instead. What we'll do is we'll plug numbers that are closer and closer to π into $f(x)$. Let's list several values of $f(x)$ as x gets closer to π from the left:

Table 2.1.3 Estimating $\lim_{x\to\pi} f(x)$

x	$f(x)$
3	$-.140$
3.1	$-.042$
3.14	$-.002$

We can even let x approach π from the other direction as well (i.e. "from the right") and $f(x)$ will still approach 0 as x gets closer and closer to π. So it looks like the limit should be 0. □

To keep ourselves from writing "the limit of $f(x)$ as x approaches some number a" over and over, let's introduce some notation: $\lim\limits_{x\to a} f(x)$.

Example 2.1.4 Limit of a piecewise function. Let

$$f(x) = \begin{cases} \sin x & \text{if } x < 0 \\ \cos x & \text{otherwise} \end{cases}.$$

Find $\lim\limits_{x\to 0} f(x)$.

Solution. If we graph this function, we see that at $x = 0$ there is a jump in the graph. In particular, if x approaches 0 from the left then $f(x)$ approaches 0, whereas if x approaches 0 from the right then $f(x)$ approaches 1. So this function does not appear to have an unambiguous limit as x approaches 0. Another way to say this: $\lim\limits_{x\to 0} f(x)$ *does not exist*. □

Example 2.1.4 shows us something very important about limits: they depend on the two different ways x can approach a number. So we introduce two new pieces of notation: the **left-hand limit** $\lim_{x\to a^-} f(x)$ will stand for the value $f(x)$ approaches (if any) as x approaches a from the left (i.e. as x increases to a), and **the right-hand limit** $\lim_{x\to a^+} f(x)$ will stand for the value $f(x)$ approaches (if any) as x approaches a from the right (i.e. as x decreases to a). In Example 2.1.4, we would say that

$$\lim_{x\to 0^-} f(x) = 0 \text{ and } \lim_{x\to 0^+} f(x) = 1.$$

At this point, we can make a rough definition for the limit of a function.

Definition 2.1.5 Limit of a Function. Let $f(x)$ be a function. Suppose that both the left-hand limit $\lim_{x\to a^-} f(x)$ and the right-hand limit $\lim_{x\to a^+} f(x)$ exist and are equal to the same number L. Then we say that the limit of $f(x)$ as x approaches a exists and is equal to L. We denote this by writing $\lim_{x\to a} f(x) = L$. ♢

Example 2.1.6 Piecewise function again. Let $f(x)$ be given by

$$f(x) = \begin{cases} 0 & -3 \le x \le -1 \\ x^2 & -1 < x < \frac{1}{2} \\ \frac{1-x}{2} & \frac{1}{2} < x < 3. \end{cases}$$

Evaluate $\lim_{x\to -1} f(x)$ and $\lim_{x\to \frac{1}{2}} f(x)$.

Solution. If we graph $f(x)$, we get the following:

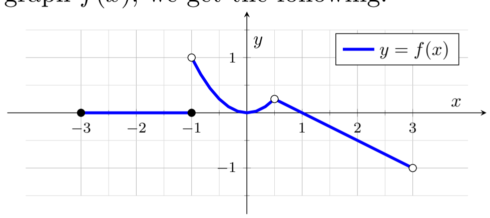


Figure 2.1.7 Graphing $f(x)$.

The graph shows us that $\lim_{x\to -1^-} f(x) = 0$, while $\lim_{x\to -1^+} f(x) = 1$. Therefore $\lim_{x\to -1} f(x)$ does not exist. On the other hand, $\lim_{x\to \frac{1}{2}} f(x)$ exists and is equal to $\frac{1}{4}$. □

It's important to note that the value of a function at a point $x = a$ is in general *completely independent* of the value of $\lim_{x\to a} f(x)$, i.e., we can't always expect $\lim_{x\to a} f(x)$ to be equal to $f(a)$. Functions for which this is true, however, are known as *continuous functions* and will be very important in Section 2.3 and beyond.

2.2 Computing Limits

We've got a handle on how to estimate limits from Section 2.1, but the process is very tedious. It requires either graphing the function in question or laboriously entering values into a calculator. So our first order of business now that we have the concept of a limit is to find an easier way to calculate it. This will be a running theme throughout the course.

The Limit Laws

In many cases of interest, we can use knowledge of simpler limits to obtain more complicated limits. We do this via the **Limit Laws**. Before we get to them, we'll state two very simple (and hopefully very believable) limits.

Proposition 2.2.1 Simple Limits. *For any value of a, the following limits hold:*

$$\lim_{x\to a} c = c$$

if c is a constant and

$$\lim_{x\to a} x = a.$$

Theorem 2.2.2 The Limit Laws. *Let c be a constant, let n be a positive whole number and let $f(x)$ and $g(x)$ be functions. Suppose that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist for some number a. Then the following rules hold:*

Table 2.2.3 The Limit Laws

1. $\lim_{x\to a}[f(x) \pm g(x)] = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x)$	2. $\lim_{x\to a} cf(x) = c\lim_{x\to a} f(x)$
3. $\lim_{x\to a}[f(x)g(x)] = \lim_{x\to a} f(x) \lim_{x\to a} g(x)$	4. $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$ *(if* $\lim_{x\to a} g(x) \neq 0$
5. $\lim_{x\to a}[f(x)^n] = [\lim_{x\to a} f(x)]^n$	6. $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)}$

Note that item six in the table above only holds (in this class...) if n is odd or if $\lim_{x\to a} f(x) \geq 0$.

Theorem 2.2.2 gives us the ability to compute a wide variety of limits.

Example 2.2.4 Limit of a rational function. Let

$$f(x) = \frac{3 - x^5 + 5x}{2x - \sqrt[4]{x}}.$$

Evaluate $\lim_{x\to 16} f(x)$.

Solution. We can evaluate $\lim_{x\to 16} f(x)$ by making use of the appropriate Limit Laws and Proposition 2.2.1:

$$\begin{aligned}
\lim_{x\to 16} \frac{3 - x^5 + 5x}{2x - \sqrt[4]{x}} &= \frac{\lim_{x\to 16}(3 - x^5 + 5x)}{\lim_{x\to 16}(2x - \sqrt[4]{x})} && \text{by Limit Law 4}\\
&= \frac{\lim_{x\to 16} 3 - \lim_{x\to 16} x^5 + \lim_{x\to 16} 5x}{\lim_{x\to 16} 2x - \lim_{x\to 16} \sqrt[4]{x}} && \text{by Limit Law 1.}\\
&= \frac{3 - [\lim_{x\to 16} x]^5 + 5\lim_{x\to 16} x}{2\lim_{x\to 16} x - \sqrt[4]{\lim_{x\to 16} x}} && \text{by Limit Laws 2, 5, and 6.}\\
&= \frac{3 - 16^5 + 80}{30}
\end{aligned}$$

□

In particular, Limit Laws 1-5 give us the following: if $f(x)$ is a polynomial or rational function, then $\lim_{x\to a} f(x) = f(a)$ as long as a is in the domain of $f(x)$. If a is not in the domain, trickery may be required.

Example 2.2.5 Trickery. Evaluate

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

Solution. First, note that we can't use the Limit Laws right away since the denominator is 0 at $x = 1$. What we need to do is use algebra to simplify the expression inside the limit:

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

Now we can use the Limit Laws to find the limit as x approaches 1, since we no longer have a divide-by-zero problem in the denominator:

$$\lim_{x\to 1} \frac{\sqrt{x} - 1}{x - 1} = \lim_{x\to 1} \frac{1}{\sqrt{x} + 1}$$

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

□

2.3 Continuity

We saw in Section 2.2 that for a function like $f(x) = x+3x^2$, we could evaluate $\lim_{x\to a} f(x)$ by simply plugging in $x = a$. In other words, $\lim_{x\to a} f(x) = f(a)$. Functions that have this property are extremely important in mathematics, so we give them a name.

Definition 2.3.1 Continuous Functions. Let $f(x)$ be a function and suppose that a is in the domain of $f(x)$. Then we say that $f(x)$ is **continuous at** a if

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

Otherwise, we say that $f(x)$ is **discontinuous at** a. We say that $f(x)$ is continuous on an interval if it is continuous at every point of an interval. Otherwise, we say that $f(x)$ is discontinuous on the interval. ♢

Definition 2.3.1 says that it is extremely easy to evaluate limits of continuous functions: just plug the value that x is approaching into the function $f(x)$. So the limit is then $f(a)$. If a function $f(x)$ is continuous on an interval, then this means that the graph of $f(x)$ has no "gaps" over this interval.

Example 2.3.2 Determining if a function is continuous. Let $f(x) = \frac{1}{x}$. Is $f(x)$ continuous on $(-\infty, \infty)$?

Solution. If we graph $f(x)$, then we see that it is discontinuous at $x = 0$. Therefore $f(x)$ is discontinuous on the interval $(-\infty, \infty)$. □

If we're dealing with a function on a (bounded) closed interval, we need to introduce some new terminology. We say that a function $f(x)$ is **continuous from the left** at $x = a$ if $\lim_{x\to a^-} f(x) = f(a)$. Similarly, we say that $f(x)$ is **continuous from the right** at $x = a$ if $\lim_{x\to a^+} f(x) = f(a)$. This is of course assuming that a is in the domain of $f(x)$. Finally, we say that $f(x)$ is continuous on the closed interval $[a, b]$ if it is continuous on (a, b), continuous from the right at a and continuous from the left at b. The main idea is still that the graph contains no gaps over this interval.

Example 2.3.3 Continuity over a closed interval. Let $f(x)$ be given by

$$f(x) = \begin{cases} 3x-1 & 1 \le x < 2 \\ 9-x^2 & 2 \le x \le 4. \end{cases}$$

Is $f(x)$ continuous over $[1, 4]$?

Solution. If we graph $f(x)$ over this interval, we get the following:

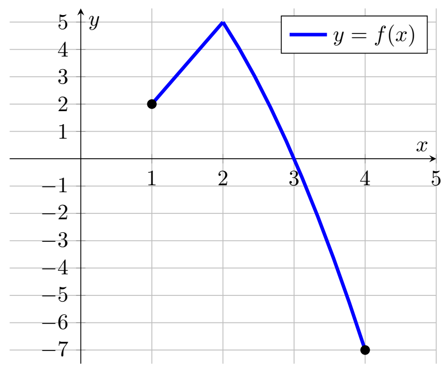


Figure 2.3.4 Graphing $f(x)$ over $[1, 4]$.

So from the graph it appears that $f(x)$ is continuous on this interval. □

Remember that we said a function is continuous over an interval if its graph has no gaps over that interval. This is a very rough explanation of continuity, but it should make the following theorem plausible.

Theorem 2.3.5 Continuous Functions. *Polynomial, rational, root and trigonometric functions are continuous at every point of their domain.*

Although it doesn't directly mention piecewise functions, Theorem 2.3.5 is still useful for determining if they are continuous. If a piecewise function is defined using any of the functions from Theorem 2.3.5, then the only points we really need to check for continuity are the the places where the function ”changes rules”.

Example 2.3.6 Another piecewise function. Over what intervals is the function $g(s)$ given by

$$g(s) = \begin{cases} s^2 & s < -1 \\ -5\cos(\pi s) & -1 \leq s \leq 1 \\ 5 & s > 1 \end{cases}$$

continuous?

Solution. We need to check continuity at $s = -1$ and $s = 1$. At $s = -1$, we need to make sure that $\lim_{s\to -1} g(s)$ exists and is equal to $g(-1)$. Since

$$\lim_{s\to -1^{-1}} g(s) = \lim_{s\to -1^{-1}} s^2 = 1$$

and

$$\lim_{s\to -1^{+}} g(s) = \lim_{s\to -1^{+}} \cos(\pi s) = -1,$$

it follows that $\lim_{s\to -1} g(s)$ does not exist. So $g(s)$ can't be continuous at $s = -1$.

On the other hand, since $\lim_{s\to 1^{-1}} g(s) = 5 = \lim_{s\to 1^{+}} g(s)$, it follows that $\lim_{s\to 1} g(s)$ exists and is equal to 5. Since $g(1)$ also equals 5, $g(s)$ is continuous at $s = 1$. So $g(s)$ must be continuous on $(-\infty, -1) \cup (-1, \infty)$. □

We can also build more complicated continuous functions out of simpler ones.

Theorem 2.3.7 Combining Continuous Functions. *Let $f(x)$ and $g(x)$ be continuous at a point a. Then the following statements are true:*

1. *$f(x) + g(x)$ is continuous at a.*

2. *$f(x)g(x)$ is continuous at a.*

3. *$\frac{f(x)}{g(x)}$ is continuous at a if $g(a) \neq 0$.*

4. *$f(g(x))$ is continuous at a if $g(a)$ is in the domain of $f(x)$.*

Example 2.3.8 Determining where functions are continuous. Let $h(t) = \sqrt{t} + \frac{t}{t-4} - \frac{t+1}{t^2-1}$. On what intervals is $h(t)$ continuous?

Solution. By Theorem 2.3.7, $h(t)$ should be continuous wherever $\sqrt{t}$, $\frac{t}{t-4}$ and $\frac{t+1}{t^2-1}$ are all defined. Since $\sqrt{t}$ is defined for $t \geq 0$, $\frac{t}{t-4}$ is defined for $t \neq 4$ and $\frac{t+1}{t^2-1}$ is defined for $t \neq \pm 1$, it follows that $h(t)$ is continuous on $[0, 1) \cup (1, 4) \cup (4, \infty)$. □

Example 2.3.9 Using continuity to evaluate a limit. Evaluate $\lim_{x \to \pi} \cos(x + \sin x)$.

Solution. Since $x, \cos x$ and $\sin x$ are all continuous, this means that $\cos(x + \sin x)$ must be continuous as well. Therefore

$$\lim_{x \to \pi} \cos(x + \sin x) = \cos(\pi + \sin \pi) = -1.$$

□

2.4 Limits Involving Infinity

Limits Involving Vertical Asymptotes

Consider the function $f(x) = \frac{1}{x}$. We know from algebra that this function has a vertical asymptote at $x = 0$, and so in particular is undefined there. However, just because it's undefined at $x = 0$ doesn't mean that we can't gather important information about the function near 0. This is because the function behaves in a very specific way as we let x approach 0. For example, if we let x approach 0 from the right, then $f(x)$ increases without bound. Similarly, $f(x)$ decreases without bound as x approaches 0 from the left.

Even though $f(x)$ is not approaching a specific value as x approaches 0 from either direction, this behavior shows up often enough and is important enough that we want to introduce notation to describe it. For this function, we would say $\lim_{x \to 0^+} \frac{1}{x} = \infty$ and $\lim_{x \to 0^-} \frac{1}{x} = -\infty$.

Now consider $g(x) = \frac{1}{(x-3)^2}$. Then $\lim_{x \to 3^-} g(x) = \lim_{x \to 3^+} g(x) = \infty$ since the function increases without bound when x approaches 3 from both directions. In this case, we say that $\lim_{x \to 3} g(x) = \infty$.

It's *extremely* important to remember that the symbol ∞ is not being used to represent a number or variable that we can perform algebra on. It's a symbol indicating how a particular function is behaving at a certain point.

If $f(x)$ is a function and $\lim_{x \to a} f(x) = \pm\infty$, $\lim_{x \to a^-} f(x) = \pm\infty$ or $\lim_{x \to a^+} f(x) = \pm\infty$, then this means that the function has a vertical asymptote at $x = a$. In this course, this basically corresponds to a divide-by-zero problem.

Example 2.4.1 Infinite limit involving a rational function. Determine $\lim_{x \to 4^-} \frac{x-3}{2x-8}$.

Solution. If we try to plug in $x = 4$ into $\frac{x-3}{2x-8}$ we get $\frac{1}{0}$, which means we have run into a divide-by-zero problem. This is a good hint that the limit should be $\pm\infty$, we just need to figure out the correct sign. There are a couple ways we can do this. First, we could set up a sign chart for this function to see where it's positive and negative and then use that to see if it's increasing or decreasing without bound as $x \to 4^-$. Second, we could just plug in values of x that are closer and closer to 4 and see how the function behaves. Either way, we see that it's negative for values of x that are close to (but less than) 4. Hence $\lim_{x\to 4^-} \frac{x-3}{2x-8} = -\infty$. □

Limits at Infinity

The previous subsection involved limits of functions whose values approached $\pm\infty$. Now we look at what can happen to a function if its input values approach $\pm\infty$. First, a definition of sorts.

Definition 2.4.2 Limit at Infinity. Let $f(x)$ be a function. We say that $\lim_{x\to\infty} f(x) = L$ if $f(x)$ gets (and stays) arbitrarily close to L as x is made arbitrarily large. Similarly, we say that $\lim_{x\to-\infty} f(x) = L$ if $f(x)$ gets (and stays) arbitrarily close to L as x is made arbitrarily small. ♢

Example 2.4.3 An important limit at infinity. Let $f(x) = \frac{1}{x}$. Determine $\lim_{x\to\infty} f(x)$.

Solution. As x gets arbitrarily large, $\frac{1}{x}$ gets arbitrarily close to 0. Therefore $\lim_{x\to\infty} f(x) = 0$. □

Example 2.4.3 holds for many other reciprocal powers of x. In particular, if $n > 0$ then $\lim_{x\to\infty} \frac{1}{x^n} = \lim_{x\to-\infty} \frac{1}{x^n} = 0$.

Example 2.4.4 A limit at infinity involving cosine. Let $g(s) = \cos s^{-3}$. Compute $\lim_{s\to-\infty} g(s)$.

Solution. First, note that $g(s) = \cos \frac{1}{s^3}$. By the previous remark, we know that $\lim_{s\to-\infty} \frac{1}{s^3} = 0$. Therefore $\lim_{s\to-\infty} g(s) = \cos 0 = 1$. □

The reason we were able to find the limit in Example 2.4.4 was because of the following fact: if $\lim_{x\to a} g(x)$ exists and if $f(x)$ is continuous at $\lim_{x\to a} g(x)$, then $\lim_{x\to a} f(g(x)) = f(\lim_{x\to a} g(x))$. Basically, we can swap continuous functions with limits without causing any harm.

For limits at infinity involving powers of a variable, it is the highest power variables that determine the outcome.

Example 2.4.5 Limit at infinity of a rational function. Let

$$f(t) = \frac{100t - 3 + t^3}{5t^2 + 1} \text{ and } g(t) = \frac{1 - 2t^4}{5t^4 + 3000t}.$$

Find $\lim_{t\to\infty} f(t)$ and $\lim_{t\to-\infty} g(t)$.

Solution. Let's start with $f(t)$. To figure out what this limit should be, we could try the following. As t gets very large the t^3 term in the numerator should drown out everything else in the numerator. Similarly, the $5t^2$ term in the denominator should drown out everything else in the denominator. So for t very large, $f(t) \approx \frac{t^3}{5t^2} = \frac{1}{5}t$. Hence $f(t)$ should probably go to ∞ as t goes to ∞. To make this more precise, we'll just divide the numerator and denominator by the largest power of the denominator, t^2, and then take the limit:

$$\begin{aligned}\lim_{t\to\infty} f(t) &= \lim_{t\to\infty} \frac{100t - 3 + t^3}{5t^2 + 1} \\ &= \lim_{t\to\infty} \frac{\frac{100}{t} - \frac{3}{t^2} + t}{5 + \frac{1}{t^2}}\end{aligned}$$

$$= \infty.$$

We can find $\lim_{t\to-\infty} g(t)$ using the same idea. Just divide by the highest power in the denominator and then take the limit:

$$\begin{aligned}
\lim_{t\to-\infty} g(t) &= \lim_{t\to-\infty} \frac{\frac{1}{t^4} - 2}{5 + \frac{3000}{t^3}} \\
&= -\frac{2}{5}.
\end{aligned}$$

□

These limits at infinity have a graphical meaning as well. If $\lim_{x\to\infty} f(x)$ or $\lim_{x\to-\infty} f(x)$ exists and is equal to L, then the line $y = L$ is a horizontal asymptote of the graph of $y = f(x)$.

Example 2.4.6 Asymptotic equivalence. Two functions $f(x)$ and $g(x)$ are said to be **asymptotically equivalent**, written $f \sim g$, if the following is true:

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

Show that $\sin\frac{1}{x} \sim \frac{1}{x}$.

Solution. All we need to do is compute $\lim_{x\to\infty} \frac{\sin\frac{1}{x}}{\frac{1}{x}}$:

$$\begin{aligned}
\lim_{x\to\infty} \frac{\sin\frac{1}{x}}{\frac{1}{x}} &= \lim_{\theta\to0^+} \frac{\sin\theta}{\theta} \\
&= 1.
\end{aligned}$$

Therefore $\sin\frac{1}{x} \sim \frac{1}{x}$. □

Chapter 3

Derivatives

There are two problems that inspired the creation of calculus. The first is the tangent line problem, or rate of change problem. This problem is concerned with determining how a quantity $f(x)$ changes as x varies. We will focus on this problem and its solution, derivatives, for the next two chapters. Afterwards, we will focus on the area problem, the second problem that inspired the creation of calculus.

3.1 The Definition of the Derivative

Tangent Lines

Consider $f(x) = x^2$. If we graph this, we get a parabola. What we'd like to do is to find a way to describe how quickly this parabola is changing at a point, i.e. find the "slope" of the parabola. One way to try to deal with this is to use **secant lines**. Recall that the secant line through the points $(a, f(a))$ and $(b, f(b))$ has slope $\frac{f(b)-f(a)}{b-a}$, which is just the average rate of change of f from $x = a$ to $x = b$. If b is very close to a, then the slope of the secant line through these points should be a good approximation to the "slope" of $f(x) = x^2$ at the point $x = a$.

Example 3.1.1 Secant lines on a parabola. Let $f(x) = x^2$. Find the slope of the secant line through $(2, f(2))$ and $(x, f(x))$.

Solution. Since the slope of the secant line is the average rate of change, we get that the slope must be equal to

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

□

What Example 3.1.1 is telling us is that if $x \approx 2$, then the slope of $f(x) = x^2$ at $x = 2$ should be very close to $x + 2$, the slope of the secant line. Now we'll do a trick that shows up everywhere in calculus, and is the entire reason we introduced limits in the first place. We'll make this approximation exact by taking a limit. In particular, we'll say that the slope of $f(x) = x^2$ should be equal to

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

This is the slope of the **tangent line** to $f(x) = x^2$ at $x = 2$. Instead of an average rate of change over an interval $[2, x]$, we now have an **instantaneous rate of change** at a point $x = 2$.

Definition 3.1.2 Tangent Lines. The tangent line to a curve $y = f(x)$ through a point $(a, f(a))$ is the line passing through $(a, f(a))$ with slope given by

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

assuming this limit exists. The slope of the tangent line represents the slope of the graph of $f(x)$ at a and gives the instantaneous rate of change of $f(x)$ at $x = a$. ◊

Example 3.1.3 Tangent line to a root. Find the equation of the tangent line to $y = \sqrt{x}$ through the point $(4, 2)$.

Solution. We need two things to find the equation of a line: the slope of the line and a point on the line. Since we know the tangent line has to pass through $(4, 2)$, we just need to find the slope. The slope is given by

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

Hence the equation of the tangent line through $(4, 2)$ is

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

□

As a reminder, the slope of the tangent line represents the slope, or instantaneous rate of change, of the function.

Example 3.1.4 Velocity from position. The displacement (i.e. position) of a particle moving in a straight line is described by the function $s(t) = 3t^2 - 5t$, where t is in seconds and s is in meters. Find the velocity, or instantaneous rate of change, of the particle at $t = 3$.

Solution. The velocity is just the slope of the tangent line to $s(t)$ at $t = 3$, which we can find as follows:

$$\begin{aligned} \lim_{t\to 3} \frac{s(t) - s(3)}{t - 3} &= \lim_{t\to 3} \frac{(3t^2 - 5t) - 12}{t - 3} \\ &= \lim_{t\to 3} \frac{3t^2 - 5t - 12}{t - 3} \end{aligned}$$

At this step it's a little unclear where to go next, so we'll try long division. If we do so, we get

$$\frac{3t^2 - 5t - 12}{t - 3} = 3t + 4.$$

Hence the velocity must be

$$\lim_{t\to 3}(3t + 4) = 13$$

meters per second. □

Example 3.1.4 was a little tricky because we needed to compute $\lim_{t\to 3} \frac{3t^2-5t-12}{t-3}$, and it was unclear how to simplify this at first. This stemmed in large part from how we computed the velocity in the first place, using the formula

$$\lim_{t\to 3} \frac{s(t) - s(3)}{t - 3}$$

or more generally

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

We want to rewrite this formula to make it a little easier to work with in certain cases. We'll do this by making the denominator easier to handle. In particular, set $x-a=h$. Then

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

Either formula can be used to compute the slope of the tangent line.

Example 3.1.5 Velocity revisited. Let $s(t)$ be given as in Example 3.1.4. Find the velocity at $t=3$.

Solution. We know that the velocity should be 13, but we'll try to find it again using our new formula. So the velocity should also be

$$\begin{aligned}\lim_{h\to 0}\frac{s(3+h)-s(3)}{h}&=\lim_{h\to 0}\frac{3(9+6h+h^2)-5(3+h)-12}{h}\\&=\lim_{h\to 0}\frac{13h+3h^2}{h}\\&=\lim_{h\to 0}(13+3h)\\&=13.\end{aligned}$$

□

Typically, if $f(x)-f(a)$ is easy to factor in terms of $x-a$ we'll want to use the first formula we had for computing rates of change. Otherwise, we'll stick with the new formula involving h.

The Derivative

Suppose we go back to Example 3.1.4 one more time, but now we want to find the velocity of $s(t)$ at an arbitrary number a. Then we could still use our limit formulas, which would give us

$$\begin{aligned}\lim_{h\to 0}\frac{s(a+h)-s(a)}{h}&=\lim_{h\to 0}\frac{3(a^2+2ah+h^2)-5(a+h)-3a^2+5a}{h}\\&=\lim_{h\to 0}\frac{6ah+3h^2-5h}{h}\\&=\lim_{h\to 0}(6a+3h-5)\\&=6a-5.\end{aligned}$$

So the velocity at $t=a$ of the particle modeled by $s(t)$ is given by $6a-5$. So we can represent the velocity, or rate of change or slope of the tangent line, by a function. We call this the **derivative**.

Definition 3.1.6 Definition of the Derivative. Let $f(x)$ be a function. Then its derivative at $x=a$ is the number $f'(a)$ given by

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

assuming the limit exists. If this limit exists, we say that the function $f(x)$ is **differentiable** at a. ◊

We could also define the derivative by

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

These two limits are equivalent.

The derivative of a function $f(x)$ at a point a represents two things: the slope of the tangent line to $f(x)$ at a and the instantaneous rate of change of $f(x)$ at a.

Example 3.1.7 Slope of the sine function. Let $f(x) = \sin x$. Find its slope at 0.

Solution. The slope at 0 is exactly $f'(0)$, which is

$$\lim_{h \to 0} \frac{\sin(0 + h) - \sin(0)}{h} = \lim_{h \to 0} \frac{\sin h}{h} = 1.$$

□

Example 3.1.8 Tangent line of the sine function. Find the equation of the tangent line to $f(x) = \sin x$ at 0.

Solution. The tangent line must pass through $(0, 0)$ and must have slope $f'(0) = 1$, so its equation is

$$y - 0 = 1(x - 0)$$

or just $y = x$. □

3.2 The Derivative as a Function

The Derivative Function

There's no reason we can't look at an arbitrary value for a in the definition of $f'(a)$ given in Definition 3.1.6. If we do this, we can define the *derivative function*.

Definition 3.2.1 The Derivative Function. Let $f(x)$ be a function. The **derivative function**, or more simply **derivative**, of $f(x)$ is the function $f'(x)$ defined by

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

assuming this limit exists. This is also often denoted by $\frac{d}{dx}(f)$ or $\frac{df}{dx}$. If this limit exists for all x in some interval I, we say that f is **differentiable on** I, or more simply **differentiable** if we do not wish to specify the interval. ♢

Example 3.2.2 Computing a derivative. Compute the derivative of $f(x) = x - 3x^2$.

Solution. Using Definition 3.2.1, we have

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

$$= \lim_{h\to 0}(1 - 6x - 3h)$$
$$= 1 - 6x.$$

□

If $f(x)$ is a function, then its derivative $f'(x)$ (assuming it exists!) is a function that gives the rate of change of f at x, or equivalently the slope of the tangent line to f at x.

Example 3.2.3 Sketching a derivative. A function $g(t)$ is given by the following graph:

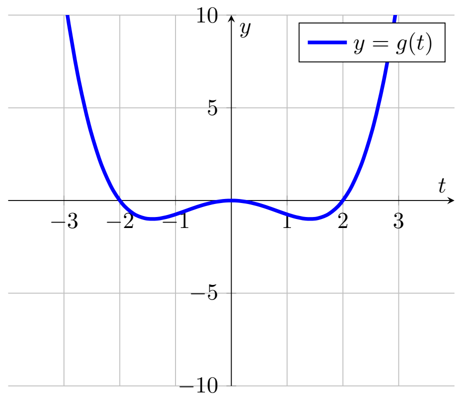


Figure 3.2.4 Graph of $g(t)$.

Sketch $\dfrac{dg}{dt}$.

Solution. Remember that $\dfrac{dg}{dt}$ represents the slope of $g(t)$, so sketching $\dfrac{dg}{dt}$ amounts to sketching the different values that the slopes of $g(t)$ can take. We can eyeball these values from Figure 3.2.4. A rough sketch of $\dfrac{dg}{dt}$, added to the original graph, may look like the following:

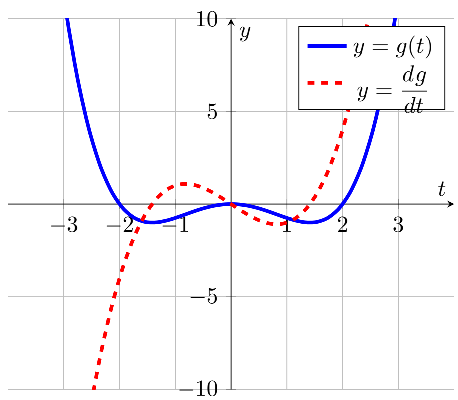


Figure 3.2.5 Graph of $g'(t)$.

□

We've mentioned before that continuous functions are functions whose graphs can be drawn without lifting your pencil off of the page. Likewise, differentiable functions are functions whose graphs can be drawn "smoothly", without any sudden movements or cusps, and without drawing a vertical tangent line. If we think about these two concepts, we may suspect that a differentiable function is also continuous. If we can draw a graph smoothly, we certainly can't lift our pencil off the page to draw it. The next theorem makes this precise.

Theorem 3.2.6 Differentiable Functions Are Continuous. *Let $f(x)$ be a function that is differentiable at $x = a$. Then $f(x)$ is continuous at $x = a$.*

Proof. We need to show that $\lim_{x\to a} f(x)$ exists and is equal to $f(a)$. To do this, we'll start by considering (somewhat counterintuitively) $\lim_{x\to a}[f(x) - f(a)]$:

$$\begin{aligned}\lim_{x\to a}[f(x)-f(a)] &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}(x-a)\\ &= \lim_{x\to a} f'(a)(x-a)\\ &= 0.\end{aligned}$$

Note that we are using our alternate definition of the derivative here.

Now we can prove that $\lim_{x\to a} f(x) = f(a)$ as follows:

$$\begin{aligned}\lim_{x\to a} f(x) &= \lim_{x\to a}[f(x)-f(a)+f(a)]\\ &= \lim_{x\to a}[f(x)-f(a)] + \lim_{x\to a} f(a)\\ &= 0 + f(a).\end{aligned}$$

So $\lim_{x\to a} f(x) = f(a)$, which means that $f(x)$ is continuous at a. ■

At this point we might think that a continuous function should also be differentiable, but this is not the case.

Example 3.2.7 A continuous function that is not differentiable at a point. Let $f(x) = |x|$. Show that f is *not* differentiable at 0.

Solution. If we graph $f(x)$ it looks like it shouldn't be differentiable at 0 because of the cusp. We'll try to prove this mathematically by showing that the limit in Definition 3.2.1 doesn't exist if $x = 0$. First, we'll compute the left hand limit:

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

Now, the right hand limit:

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

Since these limits are different, $\lim_{h\to 0}\frac{f(0+h)-f(0)}{h}$ does not exist. Hence f is not differentiable at 0. □

You may think that a continuous function must at least be differentiable "almost everywhere" at this point. After all, how could it be possible to draw a graph without lifting your pencil off the paper that still has a cusp or a vertical tangent line *everywhere*? Most mathematicians before the 19$^{\text{th}}$ century thought this as well, until Weierstrass came up with a function, the Weierstrass function[1], that is continuous everywhere but differentiable *nowhere*.

Higher Order Derivatives

Example 3.2.8 Acceleration from position. The position of some particle moving in a line is given by $s(t) = 3t - 5t^3$, where t is in seconds and s is in meters. Find $a(t)$, the acceleration of the particle at time t.

[1] en.wikipedia.org/wiki/Weierstrass_function

Solution. Acceleration is the rate of change of velocity, and velocity is the rate of change of position. So we should probably find the velocity first! Let's call it $v(t)$. We have

$$\begin{aligned}
v(t) &= \lim_{h\to 0} \frac{s(t+h)-s(t)}{h} \\
&= \lim_{h\to 0} \frac{3(t+h)-5(t+h)^3-3t+5t^3}{h} \\
&= \lim_{h\to 0} \frac{3h+5[t^3-(t+h)^3]}{h} \\
&= \lim_{h\to 0} \frac{3h+5[(-h)(t^2+t(t+h)+(t+h)^2)]}{h} \\
&= \lim_{h\to 0} (3-5[t^2+t(t+h)+(t+h)^2]) \\
&= 3-15t^2
\end{aligned}$$

Now we can get the acceleration as well:

$$a(t) = \lim_{h\to 0} \frac{v(t+h)-v(t)}{h} = -30t.$$

□

In Example 3.2.8, we had to take two derivatives of the original function $s(t)$ in order to get the acceleration $a(t)$. In other words, *acceleration is the second derivative of position.* So $a(t) = \dfrac{d}{dt}\dfrac{ds}{dt}$, which we also write as $\dfrac{d^2s}{dt^2}$ or $s''(t)$. This is an example of a **second-order derivative**. In general, we have the following definition.

Definition 3.2.9 $n^{\mathbf{th}}$**-order Derivatives.** Let $f(x)$ be a function. The $n^{\mathbf{th}}$**-order derivative** of $f(x)$ is the function obtained by differentiating $f(x)$ n times. This function is denoted by

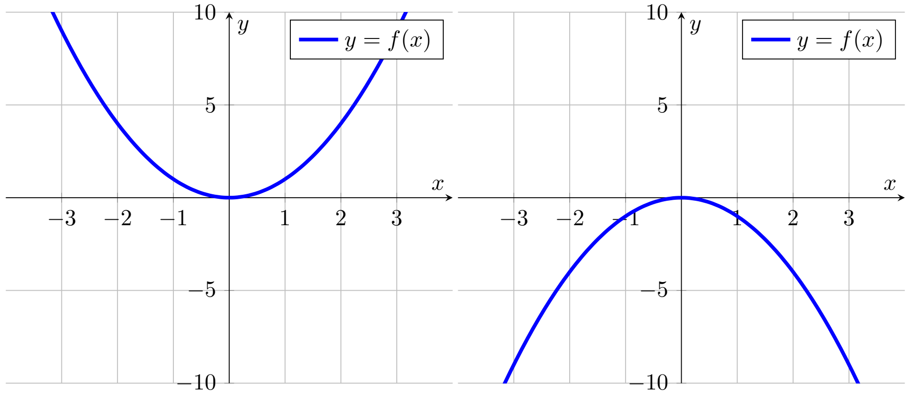
$$f^{(n)}(x) \quad \text{or} \quad \frac{d^nf}{dx^n}.$$

If $n = 1, 2$ or 3, we typically write $f'(x)$, $f''(x)$, and $f'''(x)$ instead of $f^{(1)}(x)$, $f^{(2)}(x)$ or $f^{(3)}(x)$. ◊

Although it gets more difficult to assign a physical or geometric significance to higher order derivatives, we can still derive meaning from the second derivative. One interpretation of the second derivative is as acceleration, as shown in Example 3.2.8, and it turns out there's a nice geometric interpretation as well. Recall that if $f(x)$ is a function then $f'(x)$ represents the slope, or rate of change, of the graph of $f(x)$ at x. Therefore $f''(x)$ represents the rate of change of the slope, i.e. how quickly the slope is increasing or decreasing. If $f''(x) > 0$ then the slope of $f(x)$ should be increasing, leading to a u-shaped graph. Conversely, if $f''(x) < 0$ then the slope of $f(x)$ should be decreasing, leading to an upside down u-shaped graph. This leads to the following definition.

Definition 3.2.10 Concavity. Let $f(x)$ be a function with second derivative $f''(x)$. We say that $f(x)$ is **concave up** (respectively, **concave down**) on an interval if $f''(x) > 0$ (respectively, $f''(x) < 0$) on that interval. ◊

So functions that are concave up on an interval tend to be u-shaped on that interval, and functions that are concave down tend to be upside down u-shaped. See Figure 3.2.11.



(a) $f''(x) > 0$ **(b)** $f''(x) < 0$

Figure 3.2.11 Concavity

3.3 Differentiation Formulas

Now we start to find methods that allow us to compute derivatives without going back to Definition 3.2.1. Perhaps the easiest rule is the **constant rule**, which just says that the derivative of a constant is 0. We'll derive more complicated rules in this section and the next two.

The Power Rule and Trigonometric Derivatives

Our first goal will be to determine a general formula for the derivative of x^n for some power n. If $n \geq 0$ is a whole number, then we can find the derivative of x^n without too much trouble. In fact, from the Definition 3.2.1 it's not too hard to show that the derivative of x is 1, the derivative of x^2 is $2x$, the derivative of x^3 is $3x^2$, and so on. This suggests our first derivative rule, the **power rule**.

Theorem 3.3.1 The Power Rule. *Let $f(x) = x^n$ where n is some real number. Then $f'(x) = nx^{n-1}$.*

Note that Theorem 3.3.1 is actually quite general: it works for all powers of x! This includes negative powers and fractional powers.

Example 3.3.2 Derivatives using the power rule. Find the derivatives of the following functions:

1. $f(x) = x^7$.
2. $g(x) = \frac{1}{\sqrt[5]{x^{11}}}$.
3. $h(x) = x^\pi$.

Solution. The derivative of $f(x)$ isn't too hard to find using the power rule, and we quickly get $f'(x) = 7x^6$. For $g(x)$, first rewrite it as $g(x) = x^{-\frac{11}{5}}$. Then $g'(x) = -\frac{11}{5}x^{-16}5$. Finally, $h'(x) = \pi x^{\pi-1}$. □

With a little bit of geometry and the squeeze theorem, we can get the derivatives of the basic trigonometric functions $\sin x$ and $\cos x$.

Theorem 3.3.3 Derivatives of Sine and Cosine. *Let x be in radians.*

Then

$$\frac{d}{dx}\sin x = \cos x \quad \text{and} \quad \frac{d}{dx}\cos x = -\sin x.$$

Note that if x is in degrees instead of radians these formulas don't work. Instead, they become

$$\frac{d}{dx}\sin x = \frac{\pi}{180}\cos x \quad \text{and} \quad \frac{d}{dx}\cos x = -\frac{\pi}{180}\sin x.$$

Example 3.3.4 Concavity of the sine function. On which intervals is $f(x) = \sin x$ concave up?

Solution. We need to find where $f''(x)$ is positive. Since $f''(x) = -\sin x$, this means we need to figure out where $\sin x$ is *negative*. If we go back to the unit circle definition of sine, then we can see that $\sin x < 0$ on the following intervals:

$$\ldots, (-3\pi, -2\pi), (-\pi, 0), (\pi, 2\pi), (3\pi, 4\pi), \ldots$$

So $\sin x$ is concave up on every open interval of the form $((2k-1)\pi, 2k\pi)$ where k is some integer. □

Derivatives of Sums and Constant Multiples

Now that we have derivative formulas for some basic functions, we want to extend these to more complicated functions. For this section we'll look at what happens when we multiply a function by a constant or add it to another function. In the next two sections we'll consider more advanced rules.

Theorem 3.3.5 Constant Multiple Rule. *Let $f(x)$ be differentiable function and let c be some constant. Then $\frac{d}{dx}[cf(x)] = cf'(x)$.*

Proof. To prove this, we go back to Definition 3.2.1:

$$\begin{aligned}\frac{d}{dx}[cf(x)] &= \lim_{h\to 0}\frac{cf(x+h) - cf(x)}{h}\\ &= c\lim_{h\to 0}\frac{f(x+h) - f(x)}{h}\\ &= cf'(x).\end{aligned}$$

Hence the derivative of $cf(x)$ is $cf'(x)$. ■

We can find the derivative of the sum of two functions just as easily.

Theorem 3.3.6 Sum Rule. *Let $f(x)$ and $g(x)$ be two differentiable functions. Then $[f(x) + g(x)]' = f'(x) + g'(x)$.*

Product Rule and Quotient Rule

Many functions can be written in the form $f(x)g(x)$, where $f(x), g(x)$ may each have previously known derivatives. What we want to do now is to find a way to get the derivative of $f(x)g(x)$ from the derivatives of $f(x)$ and $g(x)$. We do this using the **product rule**.

Theorem 3.3.7 The Product Rule. *Let $f(x)$ and $g(x)$ be differentiable functions. Then*

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

Proof. We prove this using the definition of the derivative:

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

■

Example 3.3.8 Using the product rule. Let $J(v) = (v^{10} - 2\sin v)(\cos v + \frac{1}{\sqrt[3]{x^4}})$. Find $J'(v)$.

Solution. We could foil this out and take derivatives, but it will be easier to use the product rule. □

The Quotient Rule

Now that we know how to differentiate products, we move on to quotients.

Theorem 3.3.9 The Quotient Rule. *Let $f(x)$ and $g(x)$ be differentiable functions. Then*

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

wherever $g(x) \neq 0$.

Example 3.3.10 Derivative of tangent. Let $f(x) = \tan x$. Find $f'(x)$.

Solution. Since $\tan x = \frac{\sin x}{\cos x}$, then we can apply the quotient rule to get the derivative of $\tan x$:

$$\frac{d}{dx}\tan x = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \sec^2 x.$$

□

The derivatives for $\sec x, \csc x$ and $\cot x$ may also be computed using the quotient rule and the facts that $\dfrac{d}{dx}\sin x = \cos x$ and $\dfrac{d}{dx}\cos x = -\sin x$.

3.4 The Chain Rule

At this point, we can take derivatives of sums, differences, products and quotients of functions. However, these rules aren't very useful for differentiating functions like $f(x) = (3x^2 + \sin x)^{50}$. We could technically evaluate $f'(x)$ using these rules but it would be an awful way to spend your weekend. But if we make the substitution $u = 3x^2 + \sin x$, then we can rewrite $f(x)$ as $f(u) = u^{50}$, which is *much* easier to differentiate: $\dfrac{df}{du} = 50u^{49}$. At this step we might be tempted to say that $f'(x) = 50u^{49} = 50(3x^2 + \sin x)^{49}$, but this isn't quite right. To get the actual derivative we need to consider how the new variable u depends on x as well. The **Chain Rule** is what we need.

Theorem 3.4.1 The Chain Rule. *Let $y = f(u)$ and $u = g(x)$ be differentiable functions. Then*

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

Equivalently, if we set $F(x) = f(g(x))$ we have

$$F'(x) = f'(g(x))g'(x).$$

Example 3.4.2 Using the chain rule. Let $f(x) = (3x^2 + \sin x)^{50}$. Find $f'(x)$.

Solution. We'll try the same trick we used before and we'll set $u = 3x^2+\sin x$. Then the chain rule says that

$$\begin{aligned} f'(x) &= \frac{df}{du}\frac{du}{dx} \\ &= (50u^{49})(6x + \cos x) \\ &= 50(3x^2 + \sin x)^{49}(6x + \cos x). \end{aligned}$$

□

As the last example highlighted, we can use the chain rule in combination with any of the other derivative rules we know if the function we're differentiating is complicated.

Example 3.4.3 Combining rules. Let $f(s) = \sqrt{\frac{s^2-1}{s^2+1}}$. Find $f'(s)$.

Solution. We'll let $u = \frac{s^2-1}{s^2+1}$ stand in for the "inside function." Then we have $f(u) = \sqrt{u}$ and so

$$\begin{aligned} \frac{df}{ds} &= \frac{df}{du}\frac{du}{ds} \\ &= \frac{1}{2}u^{-\frac{1}{2}}\frac{2s(s^2+1) - 2s(s^2-1)}{(s^2+1)^2} \\ &= \frac{1}{2}\left(\frac{s^2-1}{s^2+1}\right)^{-\frac{1}{2}}\frac{2s(s^2+1) - 2s(s^2-1)}{(s^2+1)^2} \end{aligned}$$

□

Example 3.4.4 Chain rule within a chain rule. Find the slope of $g(x) = \sin^4(\cos^3 x)$ at $x = \pi$.

Solution. We need to compute $g'(\pi)$. First, note that $g(x) = [\sin(\cos^3 x)]^4$, so let $u = \sin(\cos^3 x)$. Now we could try to use the chain rule right here but this would require finding $\dfrac{du}{dx}$, and u is itself a complicated function of x. So let $v = [\cos x]^3$, and finally let $w = \cos x$. Then we can say that

$$\begin{aligned} g'(x) &= \frac{dg}{du}\frac{du}{dv}\frac{dv}{dw}\frac{dw}{dx} \\ &= \left(\frac{d}{du}u^4\right)\left(\frac{d}{dv}\sin v\right)\left(\frac{d}{dw}w^3\right)\frac{d}{dx}\cos x \\ &= (4u^3)(\cos v)(3w^2)(-\sin x). \end{aligned}$$

We could plug in what u, v, w are in terms of x and then plug in $x = \pi$, but it's easier to just find u, v, w at $x = \pi$ and enter these values into the above. At $x = \pi$ we have $u = \sin(-1), v = -1$ and $w = -1$, so

$$g'(\pi) = 4(\sin(-1))^3\cos(-1)(3)\cdot 0 = 0.$$

□

3.5 Implicit Differentiation

Example 3.5.1 Derivative of an implicit function. Consider the curve given by the equation $x^3 - y^2 + \sin y = 3$. Find the slope of this curve at the point $(\sqrt[3]{3}, 0)$.

Solution. We could try to solve for y and then differentiate that to find the slope, but we have a slight problem: it's impossible, at least in terms of "elementary functions". However, we can still use the chain rule to find y', at least in terms of x and y. We just need to remember that y is a function of x. If we differentiate $x^3 - y^2 + \sin y = 3$ with respect to x, we get

$$3x^2 - 2y\frac{dy}{dx} + \cos y\frac{dy}{dx} = 0$$

and so

$$\frac{dy}{dx} = -\frac{3x^2}{\cos y - 2y}.$$

So the slope of the curve at $(\sqrt[3]{3}, 0)$ is just $-3\sqrt[3]{3}^2$. □

The method we used to get $\frac{dy}{dx}$ in Example 3.5.1 is called **implicit differentiation**. It's extremely useful if we want to solve for $\frac{dy}{dx}$ without first solving for y. Even if we can solve for y without too much trouble, it's often easier to find $\frac{dy}{dx}$ implicitly as the next example shows.

Example 3.5.2 Implicit differentiation to save algebra. Let $f(x) = \sqrt{1-x^2}$. Find $f'(x)$.

Solution. If we let $y = f(x)$, then $x^2 + y^2 = 1$. Then

$$2x + 2yy' = 0$$

which means that

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

□

We must also be aware of when to use appropriate derivative rules when doing implicit differentiation.

Example 3.5.3 Chain and quotient rule. Suppose y is defined implicitly by $\tan(x-y) = \frac{y}{1+x^2}$. Find y'.

Solution. We start by taking the derivative with respect to x of each side of the equation:

$$\begin{aligned}\frac{d}{dx}\tan(x-y) &= (1-y')\sec^2(x-y)\\ \frac{d}{dx}\frac{y}{1+x^2} &= \frac{y'(1+x^2) - 2xy}{(1+x^2)^2}.\end{aligned}$$

Therefore

$$(1-y')\sec^2(x-y) = \frac{y'(1+x^2) - 2xy}{(1+x^2)^2},$$

and we can solve this for y':

$$(1+x^2)^2\sec^2(x-y) + 2xy = y'(1+x^2) + (1+x^2)^2 y'\sec^2(x-y)$$

or just

$$y' = \frac{(1+x^2)^2 \sec^2(x-y) + 2xy}{(1+x^2) + (1+x^2)^2 \sec^2(x-y)}$$

□

Example 3.5.4 A differential equation. Let $P(t)$ denote the population of the United States, where P is in millions and t is the number of years after 1990. Then the growth of $P(t)$ can be modeled by the **differential equation**

$$P' = \frac{1}{23500} P(532 - P).$$

According to this model, it's not too hard to see that P' should be positive given the current population of the US, so the model predicts the population to increase. Is this rate of growth increasing or decreasing?

Solution. We need to find P'', which is the rate of change of P'. This means we need to differentiate both sides of the differential equation $P' = \frac{1}{23500} P(532 - P)$ with respect to t:

$$P'' = \frac{1}{23500} P'(532 - P) - \frac{1}{23500} P(P').$$

The current population is about 323.1 million, so we can take $P = 323.1$, which also gives

$$P' = \frac{1}{23500}(323.1)(532 - 323.1) = 2.9,$$

and so

$$P'' = \frac{1}{23500}(2.9)(532 - 323.1) - \frac{1}{23500}(323.1)(2.9) < 0.$$

Hence it appears that the rate of population increase is itself decreasing, which implies that the population growth of the US is slowing down. □

3.6 Related Rates

Example 3.6.1 Changing volume on a sphere. The radius of a sphere is increasing at a rate of $4\,\frac{\mathrm{mm}}{\mathrm{s}}$. How fast is the volume increasing when the radius is 13 mm?

Solution. First, let's assign names to all of the changing quantities in this problem:

$$V = \text{volume}$$
$$r = \text{radius}$$

Note that we're considering V and r to be functions of time, i.e. $V = V(t)$ and $r = r(t)$, where t is in seconds. Then we're given that $r' = 4$ and we need to find V' when $r = 13$. To do this, we had better find some relationship between V and r. We can find this by looking at the volume formula for a sphere:

$$V = \frac{4}{3}\pi r^3.$$

This equation relates r and V, and if we take the derivatives of both sides with respect to time t (using implicit differentiation) then we get an equation relating V' and r'. So let's do that:

$$V' = 4\pi r^2 r'$$

Now we can plug in our given information to get

$$V' = 4\pi(13^2)(4) = 2704\pi.$$

So the volume is increasing at a rate of $2704\pi \, \frac{\text{mm}^3}{\text{s}}$. □

Note that we *never* found what $V(t)$ and $r(t)$ were in Example 3.6.1, but we didn't need to. All we needed to find was a relationship between the two changing quantities, i.e. the two derivatives, to answer the question.

Example 3.6.2 A tank problem. Water is leaking out of a conical tank at a rate of $10000 \, \frac{\text{cm}}{\text{min}}$ while water is being pumped into the tank at some (unknown) constant rate. The tank has a height of 6 m and the diameter of the top is 4 m. If the water level is rising at a rate of $20 \, \frac{\text{cm}}{\text{min}}$ when the height of the water is 2 m, what is the rate at which water is being poured into the tank?

Solution. We have a *lot* of information to process here, so we'll take things a step at a time. The changing quantities are

$$\begin{aligned} h(t) &= \text{height of water} \\ r(t) &= \text{radius of water} \\ V(t) &= \text{volume of water} \end{aligned}$$

where t is in minutes. We need to find the rate at which water is entering the tank, so let's call this mystery number C. Then all we know about C is that it's related to the rate that the volume is changing. In particular, we should have $V' = C - 10000$. So to find C we need to find V', which means we need to set up a relationship between V and the other changing quantities to determine how exactly V is changing, using the fact that $h' = 20$ when $h = 200$ (after converting to centimeters).

Using the fact that the water is in a conical tank, we can use the formula for the volume of a cone to say that $V = \frac{1}{3}\pi r^2 h$. If we differentiate both sides with respect to t, then we get

$$V' = \frac{1}{3}\pi(2rr'h + r^2h').$$

Unfortunately, we don't have any information on r or r' that we can use, just information on h and h'. So we need to get r in terms of h. By similar triangles, we can say that $\frac{r}{h} = \frac{200}{600} = \frac{1}{3}$, and so $r = \frac{h}{3}$ and likewise $r' = \frac{h'}{3}$. Hence

$$V' = \frac{1}{3}\pi(2rr'h + r^2h') = \frac{1}{3}\pi\left(2\frac{h^2h'}{9} + \frac{h^2h'}{9}\right)$$

which boils down to

$$V' = \frac{h^2h'}{9}\pi = \frac{800000\pi}{9}.$$

Now, *finally*, we can answer the original question. The rate that water is flowing into the tank is

$$C = V' + 10000 = \frac{800000\pi}{9} + 10000,$$

or just $289252.68 \, \frac{\text{cm}^3}{\text{min}}$. In terms of meters this is $.29 \, \frac{\text{m}^3}{\text{min}}$, which perhaps looks a bit more reasonable. □

3.7 Linear Approximations

Suppose we graph the tangent line to $y = \sqrt{x}$ at $x = 4$. The equation of this line will be

$$y - 2 = \frac{x-4}{4} \quad \text{or} \quad y = 2 + \frac{x-4}{4}.$$

If we look at the tangent line next to the graph, we'll see that the line is very close to the graph of $y = \sqrt{x}$ if x is very close to 4. So we can use the equation of the tangent line to approximate $y = \sqrt{x}$ when x is near 4. For example, we can say that

$$\sqrt{4.2} \approx 2 + \frac{4.2-4}{4} = 2.05.$$

We call this the **linear approximation** of $\sqrt{x}$ at 4.

Definition 3.7.1 Linear Approximation. Let $f(x)$ be a function that is differentiable at $x = a$. The linear approximation (or **linearization**) of f at a is the function $L(x)$ given by

$$L(x) = f(a) + f'(a)(x - a).$$

$\diamondsuit$

The formula of the linear approximation to a function f at a is nothing more than the equation of the tangent line to f at a.

Example 3.7.2 Linear approximation of sine. Find the linear approximation of $f(\theta) = \sin\theta$ at $\theta = 0$ and use this to estimate $\sin(-.02)$.

Solution. The linear approximation is given by

$$L(\theta) = f(0) + f'(0)(\theta - 0) = \sin 0 + \cos 0(\theta - 0),$$

so $L(\theta) = \theta$. Therefore

$$\sin(-.02) \approx -.02$$

since $-.02$ is very close to 0. $\square$

Example 3.7.3 Estimating a tangent. Estimate $\tan(\pi + \frac{3}{49})$.

Solution. We want to estimate this using a linear approximation, and we want to make sure the linear approximation is easy to set up. This means we want to base the linear approximation at a value of a that both tangent and its derivative are relatively easy to compute at, and is also close to $\pi + \frac{3}{49}$. So we'll pick $a = \pi$ and find the linear approximation to $\tan x$ at $a = \pi$: Set $f(x) = \tan x$. Then

$$\begin{aligned} L(x) &= f(a) + f'(a)(x - a) \\ &= \tan\pi + \sec^2(\pi)(x - \pi) \\ &= x - \pi. \end{aligned}$$

So

$$\tan(\pi + \frac{3}{49}) \approx (\pi + \frac{3}{49}) - \pi = \frac{3}{49}.$$

$\square$

Example 3.7.4 Estimating the population of the United States in 2020. Use the differential equation from Example 3.5.4 and the fact that the current population of the US is about 323.1 million to estimate the population of the US in 2020.

Solution. Recall that $P(t)$ in Example 3.5.4 represented the population of the US (in millions) t years after 1990. To estimate the population in 2020, we want to find the linearization of P at $t = 27$. This is given by

$$L(t) = P(27) + P'(27)(t - 27).$$

Since $P(27) = 323.1$ and

$$P'(27) = \frac{1}{23500} P(532 - P) = \frac{1}{23500}(323.1)(532 - 323.1) = 2.9,$$

we have

$$L(t) = 323.1 + 2.9(t - 27).$$

Hence the population of the US in 2020 should be

$$L(30) = 323.1 + 2.9 \cdot 3 = 331.8,$$

or about 331.8 million people. □

Chapter 4

Inverse Functions

Recall that a function $f(x)$ is **invertible** if it's one-to-one. In this case, there's an inverse function $f^{-1}(x)$ that undoes $f(x)$. In symbols, $f^{-1}(f(x)) = x$ and $f(f^{-1}(y)) = y$.

4.1 Exponential Functions

Many changing quantities can be modeled by using **exponential functions**, which are functions of the form $f(x) = a^x$ where $a > 0$. Exponential functions have several important properties which we list below.

Theorem 4.1.1 Properties of Exponential Functions. *Let $a > 0$ and x, y be real numbers. Then*

- $a^0 = 1$ *and* $a^1 = a$.
- $a^{-x} = \frac{1}{a^x}$.
- $a^{x+y} = a^x a^y$ *and* $a^{x-y} = \frac{a^x}{a^y}$.
- $a^{x/y} = \sqrt[y]{a^x}$ *is* $y > 0$.
- $(a^x)^y = a^{xy}$.

Finally, if $b > 0$ as well, then $(ab)^x = a^x b^x$.

Example 4.1.2 Limit of an exponential function. Let $f(x) = 2^{-x}$. Find $\lim_{x\to\infty} f(x)$.

Solution. First, note that

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

So as $x \to \infty$, the denominator $2^x \to \infty$ as well. Hence the fraction as a whole should decrease to 0, and so

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

□

In general, we have the following:

$$a > 1 \Rightarrow \lim_{x\to\infty} a^x = \infty \quad \text{and} \lim_{x\to-\infty} a^x = 0$$

$$a < 0 \Rightarrow \lim_{x\to\infty} a^x = 0 \quad \text{and} \quad \lim_{x\to-\infty} a^x = \infty$$

Every exponential function $f(x) = a^x$ is continuous everywhere. In fact, they are differentiable everywhere, though we'll see more about that later. We can also say the following: if $a > 1$ then a^x is an increasing function, and if $0 < a < 1$ then a^x is a decreasing function. In neither case will a^x have any local maxima or minima.

Of all the different exponential functions, the most important one is the so-called **natural exponential function**, denoted e^x where

$$e = 2.71828...$$

This is the unique exponential function by the property that its derivative at 0 is 1. In fact, as we will see later e^x is its own derivative.

Example 4.1.3 Exponential derivative. Using the fact that $\frac{d}{dx}e^x = e^x$, compute $\frac{d}{dx}e^{5x-x^4+\cos x}$.

Solution. We'll have to use the chain rule. If we do so, we get

$$\frac{d}{dx}e^{5x-x^4+\cos x} = (5 - 4x^3 - \sin x)e^{5x-x^4+\cos x}.$$

□

Since e^x is its own derivative, this makes computing derivatives of functions of the form $e^{f(x)}$ relatively straightforward:

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

Technically, we have worked with the natural exponential function before in this class. Recall from Example 3.5.4 that if $P(t)$ represents the population of the USA (in millions) t years after 1990, then we said that P satisfied the differential equation

$$\frac{dP}{dt} = \frac{1}{23500}P(532 - P).$$

If we say the population of the US in 1990 was 250 million (i.e. $P(0) = 250$), then

$$P(t) = \frac{133000}{250 + 282e^{-.023t}}.$$

Now, in Example 3.7.4 we used linear approximation to estimate the population of the US using this model in the year 2020, where we got 331.8 million people. If we use the above formula for $P(t)$, we get that the population of the US in 2020 should be about

$$P(30) = 339.8,$$

or 339.8 million people.

The Census Bureau predicted in 2014 that the population in 2020 would be 334.5 million. So our simple predictions before actually aren't too far off from what the Census Bureau is expecting. This suggests that the differential equation $P' - \frac{1}{23500}P(532 - P)$ is a reasonable model for the growth of the US.

4.2 Logarithms

Inverse Function Review

Suppose we want to solve the equation $y = f(x)$ for x. Then we can only do so unambiguously if f is an **invertible function**.

Definition 4.2.1 Invertible Functions. A function $f(x)$ is invertible if and only if it is a one-to-one function. That is, $f(x_1) = f(x_2)$ forces $x_1 = x_2$. ♢

Since a function is one-to-one if and only if its graph passes the horizontal line test, the horizontal line test also gives us a useful way to check if a function is invertible.

Example 4.2.2 Let $f(x) = e^x$, the natural exponential function introduced in Section 4.1. Then $f(x)$ is an invertible function, since its graph passes the horizontal line test. □

If a function f is invertible, then there exists an **inverse function** f^1 satisfying the following equations:

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

Essentially, f^{-1} is the rule that undoes the transformation that f applies to x. The domain of f^{-1} is the range of f and the range of f^{-1} is the domain of f. Note that $f^{-1} \neq \frac{1}{f}$!.

The graph of an inverse function is related to the graph of the original function in a very nice way. To get the graph of f^{-1} from f, just reflect the graph of f across the line $y = x$.

Example 4.2.3 Graphing the inverse of the natural exponential. Let $f(x) = e^x$. Graph $f^{-1}(x)$.

Solution. If we graph e^x and then rotate the graph across the line $y = x$, we get

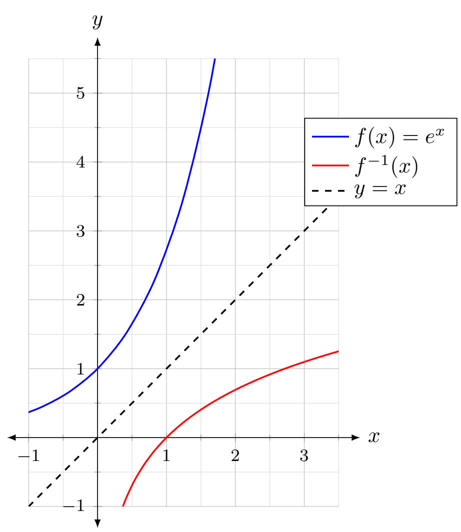


Figure 4.2.4 The inverse of e^x.

□

Calculus with Inverse Functions

There are two important things to notice about the graph of the inverse function in Example 4.2.2: it's continuous and differentiable. This is because the same was true of the original graph, and if we think about it this should be true in general as well since graphing the inverse doesn't add any new gaps or cusps to the graph.

Theorem 4.2.5 Continuity and Differentiability of Inverses. *Let $f(x)$ be a continuous (respectively, differentiable) function. Suppose that $f(x)$ is* *invertible, and has inverse $f^{-1}(x)$. Then $f^{-1}(x)$ is continuous (respectively, differentiable).*

Example 4.2.6 Finding the derivative of an inverse function. Let $f(x) = \sqrt{x-1}$ and note that f is invertible. Find the derivative of f^{-1}.

Solution. One way to do this is just start by finding f^{-1}. So we'll set $y = \sqrt{x-1}$, solve for x and then switch x and y. If we do this, we get

$$y = x^2 + 1$$

so $f^{-1}(x) = x^2 + 1$ (and note that the domain of f^{-1} is $[0, \infty)$ since this is the range of f!). Hence

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

□

The method used in Example 4.2.6 will certainly work in simple cases. But what do we do if we can't (or don't want to) find an explicit formula for the inverse function? The following formula will help us to do this.

Theorem 4.2.7 Derivative of an Inverse Function. *Let $f(x)$ be a differentiable function with inverse $f^{-1}(x)$. Assume that f^{-1} is itself differentiable. Then*

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

Proof. First, let $y = f^{-1}(x)$, so that we need to find y'. Then we can say that $f(y) = x$. Now differentiate this equation implicitly to get

$$f'(y)y' = 1$$

or just

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

■

Example 4.2.8 Exponential inverse revisited. Using the fact that the derivative of e^x is itself, find the derivative of its inverse.

Solution. Let $f(x) = e^x$. Then by Theorem 4.2.7 we have

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

So the derivative of the inverse of e^x is $\frac{1}{x}$. □

Example 4.2.9 Derivative at a point. Let $f(x) = x^3 + 3\sin x + 2\cos x$. Find $(f^{-1})'(2)$

Solution. By Theorem 4.2.7, we know that

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

By inspection, $f(0) = 2$ which means $f^{-1}(2) = 0$. Since

$$f'(x) = 3x^2 + 3\cos x - 2\sin x,$$

this gives

$$(f^{-1})'(2) = \frac{1}{f'(f^{-1}(2))} = \frac{1}{f'(0)} = \frac{1}{3}.$$

□

Logarithms

Since exponential functions are important in mathematics and its applications, their inverses are important as well. Each exponential function $f(x) = a^x$ $(a > 0)$ has an inverse function that we call the **logarithm with base** a, and denote by $f^{-1}(x) = \log_a x$. The domain of each logarithm is $(0, \infty)$ and the range is $(-\infty, \infty)$. The defining properties of the logarithm function are as follows:

$$\begin{aligned} \log_a a^x &= x \\ a^{\log_a x} &= x. \end{aligned}$$

Essentially, logarithms and exponentials cancel each other out. To put this another way, $\log_a x$ is the *exponent* needed to turn a into x.

We can say a few things about logarithms just from looking at graphs of exponentials. Suppose that $a > 1$. Then $\log_a x$ is both continuous and differentiable for all $x > 0$,

$$\begin{aligned} \log_a 1 &= 0 \\ \lim_{x\to 0^+} \log_a x &= -\infty \\ \lim_{x\to\infty} \log_a x &= \infty \end{aligned}$$

and the derivative of $\log_a x$ approaches 0 as $x \to \infty$.

The properties of the exponential listed in Theorem 4.1.1 have corresponding properties for logarithms.

Theorem 4.2.10 Properties of Logarithm Functions. *Let $a > 0$ with $a \neq 1$. Let x, y be positive real numbers, and let r be any real number. Then*

- $\log_a 1 = 0$ *and* $\log_a a = 1$.
- $\log_a(xy) = \log_a x + \log_a y$.
- $\log_a \frac{x}{y} = \log_a x - \log_a y$.
- $\log_a x^r = r \log_a x$.

An important takeaway from Theorem 4.2.10 is that logarithms turn the complicated operations of multiplication and division into the simpler operations of addition and subtraction.

Since every exponential function has a corresponding logarithm, the natural exponential function e^x has a logarithm as well. We call this inverse the **natural logarithm** and denote it by $\ln x$. Note that $\dfrac{d}{dx} \ln x = \frac{1}{x}$ by Example 4.2.8.

Example 4.2.11 Simplifying an exponential. Simplify $e^{-4\ln x}$ and $e^{x\ln 3}$.

Solution. We'll simplify by using the cancellation property $e^{\ln y} = y$. First, we need to put the entire exponent inside of the natural log:

$$e^{-4\ln x} = e^{\ln x^{-4}} = x^{-4}$$

and

$$e^{x\ln 3} = 3^x.$$

□

Example 4.2.11 shows us that *every* exponential function can be written in terms of the natural exponential: $a^x = e^{x\ln a}$. Similarly, every logarithm can be written in terms of the natural logarithm by using the change of base formula:

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

4.3 Derivatives of Exponential and Logarithmic Functions

Derivatives of Exponentials

We've already mentioned that $\frac{d}{dx}e^x = e^x$, but we've said nothing about differentiating functions such as 3^x or $(\frac{1}{\pi})^{-x}$. However, it turns out that we can get these derivatives from the derivative of e^x without too much trouble.

Theorem 4.3.1 Derivatives of Exponential Functions. *Let $a > 0$ and set $f(x) = a^x$. Then $f'(x) = a^x \ln a$.*

Proof. To prove this, we need to use what's available to us: the derivative of e^x. From Example 4.2.11 we say that $e^{x\ln 3} = 3^x$. More generally, we can also say that $e^{x\ln a} = a^x$. Hence

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

■

Example 4.3.2 Differentiating an exponential. Let $g(t) = 2.6^{3t-\sin t}$. Find $g'(t)$.

Solution. If we set $u = 3t - \sin t$, then $g(t) = g(u) = 2.6^u$. By the chain rule, $g'(t) = g'(u)\frac{du}{dt} = (\ln 2.6)2.6^u(3 - \cos t) = (\ln 2.6)2.6^{3t-\sin t}(3 - \cos t)$. □

Although we certainly want to know the derivative of a^x, or at least how to find it, most applications involving exponential functions use the natural exponential function e^x instead. The derivative of e^x is probably the most important derivative in this course.

Example 4.3.3 Solutions of a differential equation. Show that $x(t) = Ae^{-t} + Bte^{-t}$ satisfies the differential equation $x'' + 2x' + x = 0$, where A, B are arbitrary constants.

Solution. We need to show that if we compute x' and x'' and plug these expressions into the differential equation, this will simplify out to 0. Since

$$\begin{aligned} x' &= -Ae^{-t} + Be^{-t} - Bte^{-t} \\ x'' &= Ae^{-t} - Be^{-t} - (Be^{-t} - Bte^{-t}) \\ &= Ae^{-t} - 2Be^{-t} + Bte^{-t}, \end{aligned}$$

it follows that

$$\begin{aligned} x'' + 2x' + x &= (Ae^{-t} - 2Be^{-t} + Bte^{-t}) + 2(-Ae^{-t} + Be^{-t} - Bte^{-r}) + Ae^{-t} + Bte^{-t} \\ &= 0. \end{aligned}$$

□

Exponential functions often appear in the solutions of differential equations, which are themselves often viewed as mathematical models of physical systems. Hence exponential functions play a significant role in predicting physical quantities, which goes a long way towards justifying their importance.

Derivatives of Logarithms

Just as we can get the derivative of every exponential function a^x just by knowing the derivative of e^x, we can get the derivative of every logarithmic function $\log_a x$ just by knowing that $\frac{d}{dx} \ln x = \frac{1}{x}$.

Theorem 4.3.4 Derivatives of Logarithmic Functions. *Let $a > 0, a \neq 1$ and set $f(x) = \log_a x$. Then*

$$f'(x) = \frac{1}{x \ln a}.$$

Example 4.3.5 Differentiating nested logarithms. Let $f(x) = \log_2(\log_5 x)$. Find $f'(x)$.

Solution. By the chain rule, we have

$$f'(x) = \frac{1}{(\log_5 x) \ln 2} \frac{d}{dx} \log_5 x = \frac{1}{(\log_5 x) \ln 2} \frac{1}{x \ln 5},$$

□

Logarithms can be used to greatly simplify derivatives involving products, division or exponentiation using a technique known as **logarithmic differentiation**.

Example 4.3.6 A simple fraction. Let $f(x) = \frac{e^{-x} \cos x}{x^2+4x+4}$. Find $f'(x)$.

Solution. We can find $f'(x)$ without resorting to logarithms, but this would require using the product, quotient and chain rules. The algebra would be awful. So we'll using logarithms instead! Set $y = f(x)$. Then

$$\ln y = \ln e^{-x} + \ln \cos x - \ln(x^2 + 4x + 4) = -x + \ln \cos x - 2 \ln(x + 2).$$

Now we differentiate both sides implicitly to obtain

$$\frac{y'}{y} = -1 - \frac{\sin x}{\cos x} - \frac{2}{x+2}.$$

Hence

$$y' = y \left(-1 - \tan x - \frac{2}{x+2} \right) = \frac{e^{-x} \cos x}{x^2 + 4x + 4} \left(-1 - \tan x - \frac{2}{x+2} \right).$$

□

Logarithmic differentiation is useful for finding derivatives of expressions containing complicated quotients, products or powers.

Example 4.3.7 A simple exponent. Let $h(z) = z^z$. Find $h'(z)$.

Solution. We'll use logarithmic differentiation again to simplify h and remove the exponent. Set $y = h(z)$, which gives

$$\ln y = z \ln z.$$

So

$$\frac{y'}{y} = \ln z + 1,$$

which means that

$$y' = y(\ln z + 1) = z^z(\ln z + 1).$$

□

4.4 Exponential Models

Exponential Growth

Suppose we want to predict the spread of some infectious disease through a city. A reasonable, though simplistic, assumption is that the disease will spread quicker if more people are infected. In other words, we'll assume that the rate at which people are infected is proportional to the number of people infected. If we let $N(t)$ denote the number of people infected at time t, then we're basically saying that

$$\frac{dN}{dt} = kN(t)$$

for some constant $k > 0$.

This differential equation is a **mathematical model** that allows us to predict the growth of $N(t)$. We can actually use it to get an expression for $N(t)$. This differential equation is saying that $N(t)$ is a function whose derivative looks quite a bit like $N(t)$. In fact, the solution of this differential equation is $N(t) = Ce^{kt}$, where $C = N(0)$ is a constant that represents the number of people infected at time $t = 0$. This is an example of **exponential growth**, and the proportionality constant k is called the **relative growth rate**.

Example 4.4.1 Modeling an Outbreak. A game of Humans vs. Zombies breaks out at WVWC. When the game starts, there is just one zombie. After one hour, there are 3 zombies. How many zombies will there be after three hours assuming that the relative growth rate is constant?

Solution. Let $N(t)$ denote the number of zombies t hours after the game starts. Then we're given that $N(0) = 1$ and $N(1) = 3$. We need to find $N(3)$.

Since the relative growth rate is assumed to be constant, we can say that $N' = kN$ for some constant k. As we saw before, the solution of this differential equation is given by $N(t) = N(0)e^{kt} = e^{kt}$. So if we want to find $N(3)$, then we need to figure out what k is.

We can do this by using the fact that $N(1) = 3$. If we plug this into $N(t) = e^{kt}$, we get

$$3 = e^k \Rightarrow k = \ln 3.$$

So

$$N(t) = e^{(\ln 3)t} = 3^t.$$

Hence there will be $N(3) = 3^3 = 27$ zombies after three hours. □

Exponential Decay

Closely related to the concept of exponential growth is that of **exponential decay**. A quantity $M(t)$ undergoes exponential decay if it satisfies the differential equation

$$\frac{dM}{dt} = kM,$$

where $k < 0$. Every solution of this differential equation looks like $M(t) = Ce^{kt} = M(0)e^{kt}$.

We often write the decay constant k as positive, and then rewrite the ODE for $M(t)$ as $\dfrac{dM}{dt} = -kM$. The solution becomes $M(t) = M(0)e^{-kt}$. This has the effect of highlighting the negative growth rate inherent to this system.

Perhaps the most important example of exponential decay is that of radioactive decay. If $m(t)$ represents the mass of some radioactive substance, then experiments show that substance decays at a rate proportional to the amount of the substance remaining. In symbols, $\dfrac{dm}{dt} = km$ where $k < 0$. If we let $m_0 = m(0)$ denote the initial mass of the substance, then we can say that $m(t) = m_0e^{kt}$. The rate of decay is often expressed in terms of **half-life**: the amount of time it will take for precisely half of the mass to decay.

Example 4.4.2 Decay from Half-Life. A radioactive substance has a half-life of 28 days. If we start with a 50 mg sample, how much of the mass will remain after t days?

Solution. If we let $m(t)$ denote the mass of the sample at time t, and let $t = 0$ denote the first day we have the sample, then

$$m(t) = m_0e^{kt} = 50e^{kt}.$$

We still need to find k, but we can do this using the fact that the half-life is 28 days. This means that

$$m(28) = \frac{m(0)}{2} = 25.$$

So

$$25 = 50e^{28k} \Rightarrow \ln\frac{1}{2} = 28k \Rightarrow k = -\frac{\ln 2}{28}.$$

Hence

$$m(t) = 50e^{-\frac{\ln 2}{28}t} = 50 \cdot 2^{-\frac{t}{28}}.$$

□

4.5 Inverse Trigonometric Functions

Recall that $\sin x$ is not a one-to-one function since it fails the horizontal line test. Hence it has no inverse function. However, we can *restrict* the domain of $\sin x$ so that what's left over passes the horizontal line test. For example, if we define $f(x) = \sin x$ but restrict the domain to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$, then $f(x)$ *is* one-to-one and so has an inverse function. We call this function **inverse sine** or **arcsine**.

Definition 4.5.1 Inverse Sine. Let $f(x) = \sin x$ and have domain $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Then $f^{-1}(x)$ is called the inverse sine or arcsine of x, and is denoted by $\sin^{-1} x$ or $\arcsin x$. The domain is the interval $[-1, 1]$ and the range is $[-\frac{\pi}{2}, \frac{\pi}{2}]$. ◊

We can think of $\sin^{-1} x = \arcsin x$ as the angle required to turn sine into x.

Note that $\sin^{-1} x$ does *not* mean $\frac{1}{\sin x}$, which is $\csc x$. It's an unfortunate, though standard, choice of notation.

Example 4.5.2 Finding inverse sine. Find $\sin^{-1}(\frac{\sqrt{3}}{2})$.

Solution. This is the angle between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ that turns sine into $\frac{\sqrt{3}}{2}$. So $\sin^{-1}(\frac{\sqrt{3}}{2}) = \frac{\pi}{3}$. □

Example 4.5.3 Simplifying cosine and inverse sine. Simplify $\cos(\arcsin x)$, where $-1 \leq x \leq 1$.

Solution. By definition, $\sin(\arcsin x) = x$ as long as $-1 \leq x \leq 1$. So we want to try to rewrite $\cos(\arcsin x)$ to make use of this cancellation property. We can do this using the Pythagorean identity $\cos\theta = \pm\sqrt{1-\sin^2\theta}$. Hence

$$\cos(\arcsin x) = \sqrt{1-(\sin(\arcsin x))^2} = \sqrt{1-x^2}$$

whenever $-1 \leq x \leq 1$. □

Since $\sin x$ is differentiable, this means $\arcsin x$ is also differentiable.

Theorem 4.5.4 Derivative of Inverse Sine. *Let* $-1 < x < 1$. *Then* $\dfrac{d}{dx}\sin^{-1} x = \frac{1}{\sqrt{1-x^2}}$.

Proof. If we set $y = \sin^{-1} x$, then we get $\sin y = x$. Differentiating this gives $y' \cos y = 1$ or just

$$y' = \frac{1}{\cos y} = \frac{1}{\sqrt{1-x^2}}.$$

■

We can also define the **inverse cosine function** in much the same way as the inverse sine function. The idea once again is to restrict the domain of $\cos x$ to $[0,\pi]$ to get an invertible function. We call the inverse of this function the inverse cosine, or arccosine, function. We denote this function by $\cos^{-1} x$ or $\arccos x$.

Example 4.5.5 Inverse cosine value. Find $\csc(\cos^{-1}(-\frac{1}{2})$.

Solution. First, note that $\cos^{-1}(-\frac{1}{2}) = \frac{4\pi}{3}$. So

$$\csc(\cos^{-1}(-\frac{1}{2})) = \csc(\frac{4\pi}{3}) = \frac{2}{\sqrt{3}}.$$

□

Inverse cosine behaves in much the same way as inverse sine. However, this function won't be as useful to us as the inverse sine function or the next function we will look at: the inverse tangent.

If we restrict the domain of $\tan x$ to $(-\frac{\pi}{2}, \frac{\pi}{2})$, then we get an invertible function. We call the inverse of this function the **inverse tangent** or **arctangent**, and denote it by $\tan^{-1} x$ or $\arctan x$. The domain of $\arctan x$ is $(-\infty,\infty)$ and the range is $(-\frac{\pi}{2}, \frac{\pi}{2})$. $\tan^{-1} x$ can be thought of as the angle between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ that makes tangent equal to x. It satisfies the following cancellation properties: $\tan(\tan^{-1} x) = x$ for all x and $\tan^{-1}(\tan y) = y$ for all y in $(-\frac{\pi}{2}, \frac{\pi}{2})$.

Example 4.5.6 Simplifying an inverse tangent. Simplify $\sin(2\arctan x)$ using the formula $\sin 2x = 2\sin x\cos x$.

Solution. If we use the double-angle formula given, we get

$$\sin(2\arctan x) = 2\sin(\arctan x)\cos(\arctan x).$$

So we just need to find $\sin(\arctan x)$ and $\cos(\arctan x)$.

Set $\theta = \arctan x$, so that $\tan\theta = x$. Then we can find $\sin\theta$ and $\cos\theta$ using triangles, which gives

$$\sin\theta = \frac{x}{\sqrt{1+x^2}}$$
$$\cos\theta = \frac{1}{\sqrt{1+x^2}}$$

Hence

$$\begin{aligned}\sin(2\arctan x) &= \sin 2\theta \\ &= 2\sin\theta\cos\theta \\ &= \frac{2x}{1+x^2}.\end{aligned}$$

□

Example 4.5.7 Limit of inverse tangent. Determine $\lim_{x\to\infty}\tan^{-1}x$.

Solution. Recall that $\tan^{-1}x$ is the angle between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ for which tangent is equal to x. So finding $\lim_{x\to\infty}\tan^{-1}x$ is equivalent to finding what angle between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ we have to approach in order for tangent to blow up to ∞. Either from looking at a graph or by using the definition of tangent, we see that the angle we need to approach is exactly $\frac{\pi}{2}$. Hence $\lim_{x\to\infty}\tan^{-1}x = \frac{\pi}{2}$. □

Theorem 4.5.8 Derivative of Inverse Tangent. *The derivative of* $\tan^{-1}x$ *is* $\frac{1}{1+x^2}$.

Proof. We can prove this the same way we proved that $\dfrac{d}{dx}\arcsin x = \frac{1}{\sqrt{1+x^2}}$. All we need to do is to set $y = \tan^{-1}x$, and then find y' by implicit differentiation. ■

Example 4.5.9 Tangent half-angle substitution. An important substitution in integral calculus is the **tangent half-angle substitution** defined by $t = \tan\frac{x}{2}$. Use this equation to find $\dfrac{dx}{dt}$.

Solution. We can find x' implicitly, but we can also solve for x to get

$$x = 2\tan^{-1}t.$$

Therefore

$$\frac{dx}{dt} = 2\frac{d}{dt}\tan^{-1}t = \frac{2}{1+t^2}.$$

□

4.6 L'Hospital's Rule

Consider the important limit $\lim_{x\to 0}\frac{\sin x}{x}$. If we try to plug in $x = 0$, we get $\frac{0}{0}$, which is undefined. However, we can prove using geometry that the limit is 1. As another example, consider $\lim_{x\to\infty}\frac{1-x^3}{4x^3}$. Once again, if we try to plug the limit in we get an expression of the form $\frac{-\infty}{\infty}$. However, the limit is just $-\frac{1}{4}$, which we can find using algebra.

Limits of the form $\frac{0}{0}$ or $\frac{\infty}{\infty}$ are known as **indeterminate forms**. There is no restriction whatsoever on what value a limit involving an indeterminate form may take, or even that it has to exist at all.

Example 4.6.1 Different indeterminate forms. Find $\frac{\infty}{\infty}$ indeterminate forms that, respectively, evaluate to $1, 0$ and do not exist.

Solution. Let $f(x) = \frac{1-x}{3-x}$, $g(x) = \frac{1+x^4}{1+x^3}$ and $h(x) = \frac{\ln x}{\cot x}$. Then $\lim_{x\to\infty} f(x), \lim_{x\to\infty} g(x)$ and $\lim_{x\to 0^+} h(x)$ are all $\frac{\infty}{\infty}$ indeterminate forms. The first evaluates to 1, the second does not exist (it's ∞) while the third appears to be equal to 0. □

The goal of this section is to determine a method that can help us evaluate limits involving indeterminate forms. This method is called **L'Hospital's Rule**.

Theorem 4.6.2 L'Hospital's Rule. *Let $f(x)$ and $g(x)$ be differentiable functions. If $\lim_{x\to a} \frac{f(x)}{g(x)}$ is either one of the indeterminate forms $\frac{0}{0}$ or $\frac{\infty}{\infty}$, then*

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

if the limit on the right exists or is $\pm\infty$.

It's important to note that L'Hospital's Rule does not necessarily ask you to use the quotient rule!

Example 4.6.3 Using L'Hospital's Rule. Let $h(x) = \frac{\ln x}{\cot x}$. Find $\lim_{x\to 0^+} h(x)$.

Solution. We saw in Example 4.6.1 that this limit gives us the indeterminate form $\frac{\infty}{\infty}$, so L'Hospital's Rule applies. Hence

$$\begin{aligned}\lim_{x\to 0^+} \frac{\ln x}{\cot x} &= \lim_{x\to 0^+} \frac{1/x}{-\csc^2 x} \\ &= \lim_{x\to 0^+} \left(-\frac{\sin x}{x} \sin x\right) \\ &= 1 \cdot 0\end{aligned}$$

and so the limit is indeed 0. □

L'Hospital's Rule also applies for limits approaching $\pm\infty$.

Example 4.6.4 Exponential and polynomial growth. Let $f(x) = ax^2 + bx + c$ and let $g(x) = e^x$, where a, b, c are arbitrary constants. Find $\lim_{x\to\infty} \frac{ax^2+bx+c}{e^x}$.

Solution. This is another example of a $\frac{\infty}{\infty}$ indeterminate form, and so L'Hospital's Rule applies. If we use it, we get

$$\begin{aligned}\lim_{x\to\infty} \frac{ax^2 + bx + c}{e^x} &= \lim_{x\to\infty} \frac{2ax + b}{e^x} \\ &= \lim_{x\to\infty} \frac{2a}{e^x} \\ &= 0.\end{aligned}$$

In other words, the exponential function grows faster than *any* quadratic function. □

Example 4.6.5 Another limit involving exponentials. Find $\lim_{t\to 1^+} \frac{9^{t-1}-4^{t-1}}{t-1}$.

Solution. This is a $\frac{0}{0}$ indeterminate form, so L'Hospital's Rule applies. We get

$$\begin{aligned}\lim_{t\to 1^+} \frac{9^{t-1} - 4^{t-1}}{t-1} &= \lim_{t\to 1^+} \frac{9^{t-1}\ln 9 - 4^{t-1}\ln 4}{1} \\ &= \ln\frac{9}{4}.\end{aligned}$$

□

L'Hospital's Rule only applies *directly* to the indeterminate forms $\frac{0}{0}$ and $\frac{\infty}{\infty}$, but these are not the only problems when L'Hospital's Rule proves useful.

Example 4.6.6 A different indeterminate form. Find $\lim_{x\to\infty} xe^{-3x-x^2}$.

Solution. If we try to evaluate the limit, we get the expression $\infty \cdot 0$. This is another indeterminate form, but is not one that we can apply L'Hospital's Rule to without doing some algebra first. We can write

$$\lim_{x\to\infty} xe^{-3x-x^2} = \lim_{x\to\infty} \frac{x}{e^{3x+x^2}}$$

which is a $\frac{\infty}{\infty}$ indeterminate form. We *can* use L'Hospital's on this expression!

If we do so, we get

$$\begin{aligned} \lim_{x\to\infty} \frac{x}{e^{3x+x^2}} &= \lim_{x\to\infty} \frac{1}{(3+2x)e^{3x+x^2}} \\ &= 0. \end{aligned}$$

So $\lim_{x\to\infty} xe^{-3x-x^2} = 0$. □

Example 4.6.6 shows that $0\cdot\infty$ indeterminate forms can be dealt with by rewriting them in the form $\frac{0}{0}$ or $\frac{\infty}{\infty}$ and then applying L'Hospital's Rule. It's trickier, but we can also deal with $\infty-\infty$ indeterminate forms by rewriting them in this way.

Example 4.6.7 $\infty-\infty$ indeterminate form. Find $\lim_{x\to\frac{\pi}{2}^{-1}}(\sec x - \tan x)$.

Solution. As $x \to \frac{\pi}{2}^-$, $\sec x \to \infty$ and $\tan x \to \infty$. So this limit is the indeterminate form $\infty - \infty$. We can't sue L'Hospital's Rule yet, but we can try to rewrite this limit as a $\frac{0}{0}$ or $\frac{\infty}{\infty}$ indeterminate form. We'll try to do this by replacing $\sec x, \tan x$ with $\sin x, \cos x$ to get

$$\begin{aligned} \lim_{x\to\frac{\pi}{2}^-} (\sec x - \tan x) &= \lim_{x\to\frac{\pi}{2}^-} \frac{1-\sin x}{\cos x} \\ &= \lim_{x\to\frac{\pi}{2}^-} \frac{-\cos x}{-\sin x} \\ &= 0. \end{aligned}$$

□

Example 4.6.8 Limit involving radicals. Find $\lim_{x\to\infty}(\sqrt{2x^2-x} - \sqrt{2x^2+1})$.

Solution. This is another $\infty-\infty$ form, so we'll try to rewrite it into a form we can use L'Hospital's Rule on. The trick here is to factor an x out so we can get a $0\cdot\infty$ form, which we've already seen how to handle:

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

□

There are three other indeterminate forms that L'Hospital's Rule can help us with (after some algebra): $1^\infty, \infty^0$ and 0^0. All of these can be found by first using logarithms.

Example 4.6.9 A natural limit. Find $\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x$.

Solution. This limit is a 1^∞ indeterminate form. We'll try taking logarithms to rewrite it as a $0\cdot\infty$ form, so set $y=\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x$. Then

$$\ln y = \lim_{x\to\infty} x \ln\left(1+\frac{1}{x}\right)$$

is a $0\cdot\infty$ form. We can use L'Hospital's by rewriting this limit at a $\frac{0}{0}$ form or $\frac{\infty}{\infty}$. Either way, we get $\ln y = 1$ and so the original limit is $y = e$. □

Chapter 5

Applications of Differentiation

If a function is differentiable, then its derivative gives us information about how that function is changing. This information is important in many different applications of calculus.

5.1 Extreme Values of Functions

It's often of interest to determine how large or small some quantity can get.

Definition 5.1.1 Absolute Extrema. Let $f(x)$ be a function defined on some domain D. Let c be in D. Then

- $f(c)$ is an **absolute maximum** of f on D if $f(c) \geq f(x)$ for all x in D.
- $f(c)$ is an **absolute minimum** of f on D if $f(c) \leq f(x)$ for all x in D.

These values, if they exist, are **extreme values**. ♢

Example 5.1.2 List the extreme values, if any, of the following functions:

1. $f(x) = e^{-x}$.
2. $g(x) = x^2$.
3. $h(x) = \sin x$.

□

Some functions may not have extreme values, but they could have values that are smaller or larger than all other values of the function "nearby".

Definition 5.1.3 Local extrema. Let $f(x)$ be a function and let c be in the domain of f. Then

- $f(c)$ is a **local maximum** if $f(c) \geq f(x)$ for all x near c.
- $f(c)$ is a **local minimum** if $f(c) \leq f(x)$ for all x near c.

Roughly, local maxima correspond to "hilltops" whereas local minima correspond to "valleys" in the graph of $f(x)$. ♢

In general, local extrema and absolute extrema can be different. However, the following theorem does provide a relationship between the two on *closed, bounded* intervals.

Theorem 5.1.4 Extreme Value Theorem. *Let $f(x)$ be a continuous function defined on the interval $[a,b]$ Then $f(x)$ has both an absolute maximum and absolute minimum on $[a,b]$. Furthermore, these values occur at either an endpoint or a local extrema.*

What Theorem 5.1.4 tells us is that we can find the extreme values of any continuous function defined on a closed, bounded interval by just checking the function at the endpoint and at its local extrema.

Example 5.1.5 Finding extreme values. Let $f(x) = (x^2-1)^2$. Given that f has local extrema at $x = -1, 0, 1$, find the extreme values of f on the interval $[\frac{1}{2}, 6]$.

Solution. Theorem 5.1.4 tells us that we can find the extreme values by checking local extrema and the endpoints of our interval. Since $x = 1$ is the only local extreme value inside of our interval, that's the only one we need to check. We have

$$f(\frac{1}{2}) = \frac{9}{16}, f(1) = 0 \quad \text{and} \quad f(6) = 35^2.$$

So the absolute minimum of f on $[\frac{1}{2}, 6]$ is 0 and the absolute maximum is 35^2. □

So if we can find where a function has local extrema, then finding absolute extrema won't be too much more difficult. Thankfully, this is relatively straightforward if the function is differentiable.

Theorem 5.1.6 Fermat's Theorem. *Let $f(x)$ be a function and let c be a local extreme value of f. If $f'(c)$ exists, then $f'(c) = 0$.*

So finding local extrema of $f(x)$ often amounts to finding where $f'(x) = 0$, i.e., where it has a root. However, we also need to worry about where $f'(x)$ doesn't exist (just think about $y = |x|$). This leads to the following definition.

Definition 5.1.7 Critical Points. Let $f(x)$ be a differentiable function. The **critical points** of $f(x)$ are the points where $f'(x) = 0$ or $f'(x)$ does not exist. ♢

Our method for finding extreme values on a closed interval will proceed as follows: find all the critical points, and then compare the values of our function at the critical points and the endpoints of the interval.

Example 5.1.8 Extreme values of tangent. Let $f(\theta) = \frac{4}{3}\theta - \tan\theta$. Find the extreme values of f on the interval $[-\frac{\pi}{3}, \frac{\pi}{3}]$.

Solution. We need to find the critical points first. We have

$$f'(\theta) = \frac{4}{3} - \sec^2\theta.$$

This is 0 if $\cos\theta = \frac{\sqrt{3}}{2}$, which occurs in our interval only if $\theta = \pm\frac{\pi}{6}$. So we need to check f at $\pm\frac{\pi}{3}, \pm\frac{\pi}{6}$. □

Example 5.1.9 Dealing with a cusp. Let $g(x) = x^{1/3} - x^{4/3}$. Find the absolute extrema of g on $[-1, 1]$.

Solution. First, find the critical points of g:

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

This is 0 at $x = \frac{1}{4}$ and undefined at 0, so our critical points are $x = 0, \frac{1}{4}$. So to find the absolute extrema we need to check g at $x = -1, 0, \frac{1}{4}, 1$. Doing so, we see that the absolute minimum value is -2 at $x = -1$ and the absolute maximum is $(\frac{1}{4})^{1/3} - (\frac{1}{4})^{4/3}$ at $x = \frac{1}{4}$. □

5.2 The Mean Value Theorem

The linear approximations we came up with in Section 3.7 are useful for estimating complicated functions with simpler, linear models. In essence, we use the derivative of a function to tell us how much to change a given value of the function in order to estimate that function. There is one problem with this approach, at least currently. We have good reason to suspect that our approximations are close to the exact values in certain circumstances, but we don't know how close. The goal of this section is to derive an estimate for the derivative that can help us to find more precise approximations.

Rolle's Theorem

We start with a theorem that is very reasonable, at least geometrically. Let $f(x)$ be a differentiable function on some interval $[a, b]$, and suppose $f(a) = f(b)$. Then this means the graph of f must "turn around" at some point in $[a, b]$, i.e., there is a local maximum or minimum contained within the interval. Combining this observation with Theorem 5.1.6 gives us **Rolle's Theorem.**

Theorem 5.2.1 Rolle's Theorem. *Let $f(x)$ be a differentiable function on $[a, b]$ and suppose $f(a) = f(b)$. Then there exists some number c in (a, b) such that $f'(c) = 0$.*

Theorem 5.2.1 is an example of an **existence theorem**. It tells us nothing about how to find the number c, or how many such c can exist. It only tells us that there is at least one number c in (a, b) for which $f'(c) = 0$. This may not seem very useful, but existence theorems can be quite powerful in mathematics.

Example 5.2.2 Rolle's Theorem and Traffic. A car enters a highway going at 45 $\frac{\text{mi}}{\text{h}}$ and leaves the highway going at the same speed. Was the car's acceleration ever 0?

Solution. The speed of the car can be represented by a velocity function $v(t)$. If we assume that the car entered the highway at time t_{in} and left at some future time t_{out}, then we know that

$$v(t_{\text{in}}) = v(t_{\text{out}}) = 45.$$

By Theorem 5.2.1, there must be some time t_1 between t_{in} and t_{out} for which $v'(t_1) = 0$. Since v' is exactly the acceleration, we know that the car had to stop accelerating somewhere on the highway. □

The Mean Value Theorem

Rolle's Theorem is powerful because it has very general conditions for its use. However, it does require the function in question to take on the same values at the endpoints of $[a, b]$, and this is a condition we'd like to try to relax. However, we'll try to be clever about this and use Rolle's Theorem to do most of the heavy lifting for us.

If we imagine graphing some differentiable function $f(x)$ on some interval $[a, b]$, but $f(a) \neq f(b)$, then we can't apply Rolle's Theorem. But maybe we can adjust it just a little bit so that we can? In particular, the only reason we can't use Theorem 5.2.1 is that $f(a) \neq f(b)$. But if we subtract the line through these points from $f(x)$, we should get a new function for which Theorem 5.2.1 applies.

The line through these points has equation

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

So define

$$g(x) = f(x) - y = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Then

$$g(a) = g(b) = 0,$$

So Theorem 5.2.1 *does* apply to this function. Hence there exists some number c between a and b for which $g'(c) = 0$.

But

$$g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a},$$

and if this is 0 then we can solve $g'(c) = 0$ for $f'(c)$ to get

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

This gives us the **Mean Value Theorem**.

Theorem 5.2.3 Mean Value Theorem. *Let $f(x)$ be differentiable on some interval $[a, b]$. Then there exists a number c in (a, b) such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

or equivalently

$$f(b) - f(a) = f'(c)(b - a).$$

The Mean Value Theorem essentially says that there is at least one point inside of (a, b) for which the slope $f'(c)$ at that point matches the "average slope" $\frac{f(b)-f(a)}{b-a}$. Like Rolle's Theorem, this is an existence theorem. However, it's slightly more general, and so is applicable in more situations. It's also useful in deriving error estimates.

Example 5.2.4 Estimating error in a linear approximation. Let $f(x) = \sqrt{x}$ and let $L(x)$ denote the linear approximation to $f(x)$ at $a = 9$. Find the largest possible error between $f(x)$ and $L(x)$ on the interval $[9, 16]$.

Solution. First, we should find $L(x)$, which from Section 3.7 is just

$$L(x) = 3 + \frac{1}{6}(x - 9).$$

The error we need to estimate is $|f(x) - L(x)|$, for x in $[9, 16]$. So we'll set $g(x) = f(x) - L(x)$ and let x be some number in $[9, 16]$. By Theorem 5.2.3, there exists some number c in $(9, x)$ such that

$$g(x) - g(9) = g'(c)(x - 9).$$

Now, $g(9) = 0$ and $g'(c) = \frac{1}{2\sqrt{c}} - \frac{1}{6}$. So the above equation becomes

$$f(x) - L(x) = \left(\frac{1}{2\sqrt{c}} - \frac{1}{6}\right)(x - 9).$$

Now, we don't know anything about c aside from the fact that it lives in $(9,16)$. However, we do have an expression for $|f(x)-L(x)|$ now:

$$\begin{aligned}|f(x)-L(x)| &= \left|\left(\frac{1}{2\sqrt{c}}-\frac{1}{6}\right)(x-9)\right| \\ &\le \left(\frac{1}{6}-\frac{1}{8}\right)7.\end{aligned}$$

So if x is in $[9,16]$, then the error $|f(x)-L(x)|$ is *at most* $\frac{7}{24}$. In other words,

$$L(x)-\frac{7}{24}\le f(x)\le L(x)+\frac{7}{24}.$$

□

Perhaps the most important example of Theorem 5.2.3 in this course will be the following theorem.

Theorem 5.2.5 Zero Derivative Theorem. *Let $f(x)$ be differentiable on (a,b). If $f'(x)=0$ for all x in (a,b), then $f(x)$ has to be a constant.*

Proof. The idea of the proof is to use Theorem 5.2.3 to show that $f(x_2)-f(x_1)=0$ for any points x_1,x_2 in (a,b). Now, by Theorem 5.2.3 we have

$$f(x_2)-f(x_1)=f'(c)(x_2-x_1)$$

for some c in (a,b). However, $f'(c)=0$ by assumption, and so $f(x_2)-f(x_1)=0$. In other words, $f(x_2)=f(x_1)$. Hence f must be constant. ■

5.3 Derivatives and Graphs

Recall that if $f(x)$ is a function and $f'(x)$ exists, then $f'(c)>0$ for some c means that f is increasing at c, while $f'(c)<0$ means that f is decreasing at c. Now, suppose that c is a critical point of $f(x)$. Then $f'(c)=0$. If f' changes sign from negative to positive, then $f'(c)$ is a local minimum. Conversely, if f' changes sign from positive to negative, then $f'(c)$ is a local maximum. This is the **first derivative test**.

Example 5.3.1 Local maxima and minima using the first derivative test. Let $f(x)=x^4e^{-x}$. Find where f is increasing, decreasing, and any local maxima or minima.

Solution. We can answer all of these questions by setting up a sign chart for

$$f'(x)=x^3e^{-x}(4-x).$$

The critical points are $x=0,4$, and $f'<0$ on $(-\infty,0)\cup(4,\infty)$ and $f'>0$ on $(0,4)$. So f is decreasing on the first set of intervals, increasing on $(0,4)$, f has a local maximum at $x=4$ and a local minimum at $x=0$. □

Example 5.3.2 First derivative test and a discontinuous function. Find any local maxima or minima of $G(x)=\frac{x^2}{x-1}$.

Solution. We need to find the critical points, which means we need to find $G'(x)$:

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

So the critical points are $x=0,1,2$. Note that $x=1$ *cannot* be a local extreme value of G since it's not in the domain of G. However, we still need to include it in our sign chart. If we do so, we find that G has a local minimum at $x = 2$ and a local maximum at $x = 0$ by the first derivative test. □

One benefit of the first derivative test is that we only need to compute first derivatives to use it. However, if a function has a second derivative, then it's often easier to use the concavity of the graph at a critical point to determine whether a critical point is a local maximum or minimum. In particular, if $f(x)$ is a function and $f''(x)$ is continuous near the critical point c (so $f'(c) = 0$), then $f(c)$ is a local minimum if $f''(c) > 0$ and $f(c)$ is a local maximum if $f''(c) < 0$. This is the **second derivative test**.

Example 5.3.3 Using the second derivative test. Find any local extrema of $h(x) = 5x^3 - 3x^5$.

Solution. First, we find the critical points:

$$h'(x) = 15x^2 - 15x^4 = 0$$

forces $x = 0, \pm 1$. Now we check these critical points in $h''(x) = 30x - 60x^3$:

$$\begin{aligned} h''(0) &= 0 \\ h''(-1) &= 30 \\ h''(1) &= -30. \end{aligned}$$

So h has a local minimum at -1 and a local maximum at 1. □

The point $x = 0$ in Example 5.3.3 is an example of an **inflection point** of $h(x)$: a place where the concavity of h changes, or equivalently a point where the second derivative changes sign.

5.4 Optimization Problems

Now that we can use derivatives to find local maxima and minima, we can solve **optimization problems**.

Example 5.4.1 Minimizing a product. Find two numbers that differ by 100 and whose product is a minimum.

Solution. Call our numbers x, y and suppose $x > y$. Then $x - y = 100$ and we need to choose x, y in such a way that xy is as small as possible. If we replace x with $100 + y$, then this means we need to minimize

$$f(y) = 100y + y^2.$$

The critical point of $f(y)$ is $y = -50$, and since f is a parabola we know that this must be the location of the absolute minimum value on the graph. So the numbers we need are $50, -50$. □

Example 5.4.2 Minimizing material costs. A box with an open top and a square base is to be built using material that costs \$5 per square meter. The box must have a volume of $32000\,\text{cm}^3$. Find the dimensions of the box that will be cheapest to build to these specifications.

Solution. There are two things we need to do here:

1. Find the quantity to be optimize.
2. Solve the resulting optimization problem.

We need to minimize costs, which means we need to minimize how much material is used to construct this box. Hence we must minimize the surface area of the box. If we let x denote the length of the base of the box and h the height of the box, then the surface area is given by

$$x^2 + 4xh.$$

Now, we can't minimize this yet because we have too many variables. However, we can solve for h in terms of x by using the volume constraint. Since the volume of the box is $x^2h = 32000$, this gives $h = \frac{32000}{x^2}$. So the quantity we need to minimize is

$$f(x) = x^2 + 4xh = x^2 + \frac{128000}{x}.$$

To minimize this, we find the critical points by solving $f'(x) = 0$. This gives, after some algebra, $x = 40$. Now we need to be careful here! We only have a critical point at this step. We need to make an argument now as to why this critical point should be an absolute minimum of $f(x)$. One way to do this is by setting up a sign chart for $f'(x)$. If we do this, then we see that $f' < 0$ for x in $(0, 40)$, while $f' > 0$ for x in $(40, \infty)$. If we think about what this tells us of the graph of $f(x)$, we see that $x = 40$ must minimize $f(x)$, at least for $x > 0$. So the dimensions that minimize the cost of building this box are $40 \times 40 \times 20$. □

Example 5.4.3 Distance on an ellipse. Find the point(s) on the ellipse $4x^2 + y^2 = 4$ that are farthest from the point $(1, 0)$.

Solution. This is an optimization problem since we're trying to maximize distance. There are two things we need to do:

1. Find the function we need to optimize.
2. Find the extrema.

The function we need to optimize is just the distance function. In particular, if (x, y) is a point on the ellipse then we need to maximize

$$d = \sqrt{(x-1)^2 + (y-0)^2}.$$

Now, we have a constraint that (x, y) must satisfy; namely, this must lie on the ellipse $4x^2 + y^2 = 4$. This means that $y^2 = 4 - 4x^2$, and if we plug this into our distance function we get

$$d = \sqrt{(x-1)^2 + (y-0)^2} = \sqrt{(x-1)^2 + 4 - 4x^2}.$$

Now here's a trick we can use: maximizing d is the same thing as maximizing d^2, but d^2 is *much* nicer to work with algebraically. So instead of maximizing d, we'll maximize the function

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

We'll start by finding it's local extrema, which means we need to find the critical points. These are the solutions of $f'(x) = 0$. Since

$$f'(x) = 2x - 2 - 8x = -6x - 2,$$

we see that the only critical point is $x = -\frac{1}{3}$. Since $f''(x) = -6 < 0$, this means that $x = -\frac{1}{3}$ gives us a local maximum of $f(x)$ by the second derivative test. In fact, we can go further: this must be an absolute maximum of $f(x)$, since $f(x)$ is always concave down (it's actually a parabola).

So the point on the ellipse $4x^2 + y^2 = 4$ farthest from $(1, 0)$ has x-coordinate equal to $-\frac{1}{3}$. This means that the corresponding y-coordinate is $y = \pm\frac{2\sqrt{5}}{3}$, so the points on the ellipse that are the farthest from $(1, 0)$ are the points $(-\frac{1}{3}, -\frac{2\sqrt{5}}{3}), (-\frac{1}{3}, \frac{2\sqrt{5}}{3})$. □

5.5 Newton's Method

Consider a differentiable function $f(x)$, and suppose we want to find a root of $f(x)$, which is a number x_0 such that $f(x_0) = 0$. In some cases this is very easy (like $f(x) = x^2 + 2x + 1$), but in others this may be more complicated (such as $f(x) = x^5 - x - 1$). So we'd like to find a way to *approximate* the root x_0.

Newton's Method begins as follows: pick some starting point, or guess, x_1. Then draw the tangent line L to the graph of $f(x)$ at x_1. Now, the x-intercept of this line may not be a root, but hopefully it'll be closer to the root. So call this point x_2. We can solve for x_2 (if $f'(x_1) \neq 0$) to get

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Now, at this point x_2 is (hopefully!) a better approximation to the root x_0 than x_1, and we can run through the same procedure to get a new point x_3.

In general, we can get a **sequence** of approximations $x_1, x_2, \ldots, x_n$ to the root x_0 by using the formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

as long as $f'(x_n) \neq 0$.

Example 5.5.1 Approximating a root. Find the third approximation given by Newton's Method for the root of $f(x) = x^5 - x - 1$, using $x_1 = -1$.

Solution. First, note that $f'(x) = 5x^4 - 1$. Then

$$\begin{aligned} x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)} \\ &= -1 - \frac{-1}{4} \\ &= 3, \end{aligned}$$

so $x_2 = 3$. Now we run through this method again to get x_3:

$$\begin{aligned} x_3 &= x_2 - \frac{f(x_2)}{f'(x_2)} \\ &= 3 - \frac{239}{404} \\ &= \frac{973}{404}. \end{aligned}$$

So $x_3 = \frac{973}{404}$. □

Example 5.5.2 Approximating π. Approximate π using Newton's method.

Solution. We need to find a function $f(x)$ for which $f(\pi) = 0$. Perhaps the easiest is $f(x) = \sin x$. Now, to use Newton's method we also need a starting guess. We'll pick $x_1 = 3$ since this is close to π. Then

$$x_2 = x_1 - \frac{\sin x_1}{\cos x_1} = 3 - \tan 3$$

so $x_2 \approx 3.142546543$. Similarly,

$$x_3 = x_2 - \tan x_2 \approx 3.141592653$$

$$x_4 \approx 3.141592654$$

Since these approximations are so close, we estimate that $\pi \approx 3.14159265$. □

Example 5.5.3 A Babylonian problem. Use Newton's method to find an algorithm for computing $\sqrt{a}$.

Solution. To use Newton's method, we need to come up with a function $f(x)$ whose root is $\sqrt{a}$. A simple choice for this is $f(x) = x^2 - a$, since $f(\sqrt{a}) = 0$. Now, if we're getting some sequence $x_1, \ldots, x_n$ from Newton's method then the next term in the sequence is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - a}{2x_n}.$$

After some algebra, we can rewrite this as

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right).$$

We can actually test this algorithm out. Say we want to approximate $\sqrt{5}$. A reasonable guess would be $x_1 = 2$, since this should be close to $\sqrt{5}$. Then we get

$$\begin{aligned} x_2 &= \frac{1}{2}\left(2 + \frac{5}{2}\right) = 2.25 \\ x_3 &= \frac{1}{2}\left(2.25 + \frac{5}{2.25}\right) \approx 2.236111111 \\ x_4 &\approx 2.236067978 \\ x_5 &\approx 2.236067977 \end{aligned}$$

So it looks like $\sqrt{5} \approx 2.23606797$, and indeed this is quite close to the actual value of $\sqrt{5} = 2.236067977\ldots$.

We can go one step further for this algorithm. If $x_1, x_2, \ldots, x_n, \ldots$ is the sequence we obtain using Newton's method for approximating $\sqrt{a}$, then as we saw earlier

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right).$$

Now, assume that $\lim_{n\to\infty} x_n$ exists and is nonzero, and call it s. Then

$$\lim_{n\to\infty} x_{n+1} = \frac{1}{2}\left(\lim_{n\to\infty} x_n + \frac{a}{\lim_{n\to\infty} x_n}\right)$$

becomes

$$s = \frac{1}{2}\left(s + \frac{a}{s}\right)$$

which we can rearrange to get

$$s^2 = \frac{1}{2}(s^2 + a),$$

or just $s^2 = a$. Hence $s = \pm\sqrt{a}$. □

5.6 Antiderivatives

Suppose we're tracking a moving object, and through experimentation we know its acceleration to be equal to $3\,\frac{\mathrm{m}}{\mathrm{s}^2}$ to the right. Say we want to find the velocity of the mass at time t. Then if we set $v(t)$ equal to the velocity and $a(t)$ equal to the acceleration, we know that

$$v'(t) = a(t) = 3.$$

Therefore $v(t)$ is a function whose derivative is 3. So one possible choice for $v(t)$ is $v(t) = 3t$. This leads us to the definition of an **antiderivative**.

Definition 5.6.1 Antiderivatives. A function $F(x)$ is called an antiderivative of the function $f(x)$ if $F'(x) = f(x)$ for all x in an interval I. ♢

Example 5.6.2 Multiple antiderivatives. Find three different antiderivatives for $f(x) = \cos x$.

Solution. An antiderivative of $\cos x$ is any function whose derivative is $\cos x$. So three antiderivatives could be

$$F_1(x) = \sin x, F_2(x) = \sin x + 4, F_3(x) = \sin x - 100.$$

□

From Example 5.6.2 we see that any function of the form $\sin x + C$ is an antiderivative of $\cos x$. In fact, this describes all possible antiderivatives of $\cos x$. This suggests the following theorem.

Theorem 5.6.3 General Antiderivatives. *Let $f(x)$ be some function with antiderivative $F(x)$ on some interval I. Then the most general antiderivative of $f(x)$ is*

$$F(x) + C$$

where C is an arbitrary constant.

Proof. It might not seem like this statement requires a proof, but it does! We can check very easily that every function of the form $F(x)+C$ is an antiderivative of $f(x)$, but how do we know every antiderivative takes this form? To prove this, let $G(x)$ be an arbitrary antiderivative of $f(x)$. Then we need to show that $G(x) = F(x) + C$, or equivalently that $G(x) - F(x) = C$. We can do this by taking the derivative of $G(x) - F(x)$ to get

$$G'(x) - F'(x) = f(x) - f(x) = 0,$$

and so by Theorem 5.2.5 we know that $G(x) - F(x)$ must be constant on I. Hence $G(x) - F(x) = C$ for some constant C, or equivalently $G(x) = F(x) + C$. Therefore *every* antiderivative of $f(x)$ can be written as $F(x) + C$. ■

It might seem superfluous, but typically when dealing with models requiring us to find antiderivatives we'll want to find the most general antiderivative somewhere along the way. Moral of the story: *don't forget to add C.*

Example 5.6.4 Finding antiderivatives. Find the most general antiderivatives of the following functions:

1. x^3.
2. $-\dfrac{1}{2}x^{\frac{5}{3}}$.
3. $\sqrt{x} - 3x^3 + x^{-2}$.
4. $\sin x + \cos(5x) + e^{4x}$.

Solution. We have the following:

1. The general antiderivative of x^3 is $\frac{1}{4}x^4 + C$.
2. The general antiderivative of $-\frac{1}{2}x^{\frac{5}{3}}$ is $-\frac{1}{2}\frac{3}{8}x^{\frac{8}{3}} + C$.
3. The general antiderivative of $\sqrt{x} - 3x^3 + x^{-2}$ is $\frac{2}{3}x^{\frac{3}{2}} - \frac{3}{4}x^4 - \frac{1}{x} + C$.
4. The general antiderivative of $\sin x + \cos(5x) + e^{4x}$ is $-\cos x + \frac{1}{5}\sin(5x) + \frac{1}{4}e^{4x} + C$.

□

Example 5.6.5 The antiderivative of $\frac{1}{x}$. Find the general antiderivative of $f(x) = \frac{1}{x}$.

Solution. We have two cases we need to consider, since the domain of $f(x)$ consists of two intervals. First, suppose that x is in $(0, \infty)$. Then we know that $\frac{d}{dx}\ln x = \frac{1}{x}$, so on this interval the most general antiderivative of $\frac{1}{x}$ is $\ln x + C_1$.

Now let x be in $(-\infty, 0)$. Then $\ln x$ isn't defined here. However, we can write

$$\frac{1}{x} = -\frac{1}{-x},$$

and if $x < 0$ then $-x > 0$ and $\ln(-x)$ is defined, and in fact

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

So the most general antiderivative of $\frac{1}{x}$ on $(-\infty, 0)$ is $\ln(-x) + C_2$. Putting this all together, we can say that the most general antiderivative of $f(x) = \frac{1}{x}$ is given by

$$F(x) = \begin{cases} \ln x + C_1 & x > 0 \\ \ln(-x) + C_2 & x < 0 \end{cases}.$$

If we know that we're only selecting values of x with the same sign, then we can just say that the antiderivative of $\frac{1}{x}$ is $\ln|x| + C$. □

Example 5.6.6 Antiderivatives involving secant. Find the general antiderivatives of $\sec^2 t$ and $-2\sec 3t\tan 3t$.

Solution. The first antiderivative isn't too hard; it's just $\tan t + C$. The second is a little more complicated but still not too bad. Since $\frac{d}{dt}\sec t = \sec t\tan t$, a good guess for an antiderivative of $-2\sec 3t\tan 3t$ would be $-2\sec 3t$. However, the derivative of this is $-6\sec 3t\tan 3t$, so we're off by a factor of 3. So we just need to divide our guess by 3 to correct for this. Hence the general antiderivative of $-2\sec 3t\tan 3t$ is $-\frac{2}{3}\sec 3t$. □

Theorem 5.6.7 Antidifferentiation Formulas. *Let $F(x)$ and $G(x)$ denote antiderivatives of $f(x)$ and $g(x)$, and let c be a constant. Then we have the following:*

Table 5.6.8 Particular antiderivatives of functions

Function	*Antiderivative*
$cf(x)$	$cF(x)$
$f(x)+g(x)$	$F(x)+G(x)$
$x^n, n \neq 1$	$\frac{x^{n+1}}{n+1}$
$\frac{1}{x}$	$\ln\lvert x\rvert$
$e^{cx}, c \neq 0$	$\frac{1}{c}e^{cx}$
$\sin cx, \cos cx, c \neq 0$	$-\frac{1}{c}\cos cx, \frac{1}{c}\sin cx$

These are by no means the only antiderivatives that you will need to deal with, but they are probably the most common.

Example 5.6.9 Falling objects. An object is dropped from a height of 100 m above sea level and falls with downward acceleration equal to $9.8\,\frac{\mathrm{m}}{\mathrm{s}^2}$. Find the height $h(t)$ of the object t seconds after it's dropped.

Solution. Take downward to be the negative direction and sea level to be $h=0$. Since acceleration is the second derivative of position, then the position (i.e. height) of the object should be the second *antiderivative* of acceleration. At this step, it's tempting to say that since $h''(t)=-9.8$, we have

$$h'(t)=-9.8t \quad \text{and} \quad h(t)=-4.9t^2.$$

And indeed, $-4.9t^2$ is a function whose second derivative is -9.8. However, we have a slight problem here. We know that $h(0)=100$ and $h'(0)=0$. If $h(t)=-4.9t^2$, then it's impossible for $h(0)$ to equal 100. The problem here is we forgot about the arbitrary constants.

To get an accurate expression for $h(t)$, we go back to $h''(t)=-9.8$. Then $h'(t)=-9.8t+C_1$ for some constant C_1. Since $h'(0)=0$, this forces $C_1=0$. So we have $h'(t)=-9.8t$. Now we antidifferentiate one more time to get $h(t)=-4.9t^2+C_2$. Since $h(0)=100$, this forces $C_2=100$. So

$$h(t)=-4.9t^2+100.$$

□

Some functions that do not have an obvious antiderivative can be simplified through algebra into a form that is perhaps more helpful.

Example 5.6.10 A tricky antiderivative. Find the most general antiderivative of

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

Solution. It's tough to think of a function whose derivative is $f(x)$, though since the denominator is $1+x^2$ it seems likely that this antiderivative will involve $\tan^{-1}x$ in some way. In order to actually find the antiderivative, we'll rewrite $f(x)$ into a more convenient form. First, note that the numerator is very close to the denominator, which means we can almost cancel it out. So we'll split up the numerator as follows:

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

We can find the antiderivative of this term by term, so the most general antiderivative of $f(x)$ is

$$F(x)=\tan^{-1}x+x+C.$$

□

At this point, it's natural to think about what an antiderivative means geometrically since the derivative of a function has such a nice interpretation. One approach to this is to think about **net change** of the antiderivative. Let $F(x)$ be an antiderivative of $f(x)$ on some interval containing a, b. The net change of $F(x)$ from $x = a$ to $x = b$ is given by

$$F(b) - F(a).$$

Now, $F'(x) = f(x)$ represents instantaneous rate of change, not net change. However, if we can somehow add these instantaneous changes from $x = a$ to $x = b$, they should accumulate until they give the net change of F from a to b. Graphically, adding these instantaneous rates of change looks like it's giving us the *area* under $f(x)$, which suggests the following geometric interpretation: *the area under of $f(x)$ from $x = a$ to $x = b$ is just the net change of an antiderivative $F(x)$ from $x = a$ to $x = b$.* Much of the next chapter is devoted to making this idea precise.

Part II

Integral Calculus

Chapter 6

Integrals

By now we have a pretty good idea of how to calculate derivatives and what they represent. In particular, we've seen that $f'(x)$ tells us how quickly $f(x)$ is changing at a given point. However, we don't yet have a good intuition for the meaning of the *antiderivative* of a function aside from a rough interpretation as the inverse of differentiation.

To help develop some intuition for what the antiderivative represents, consider a continuous function $f(x)$ defined over some interval $[a, b]$. Now let $A(t)$ denote the area between $f(x)$ and the x-axis from $x = a$ to $x = t$, where $a \leq t \leq b$:

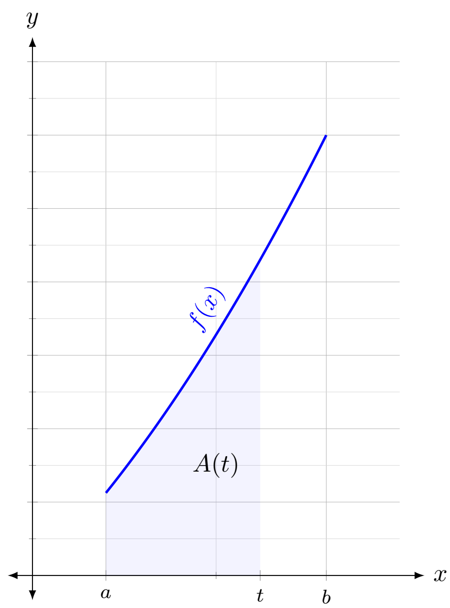


Figure 6.0.1 A graph of $f(x)$ and its corresponding area function.

Then it appears that

$$A(x + h) - A(x) \approx hf(x)$$

for small values of h. If we let $h \to 0$, then the above equation suggests that $A'(x) = f(x)$. In other words, *the area function is an antiderivative of* $f(x)$!

Going forward this will be our primary geometric interpretation of an antiderivative, but we need to address two issues to make the above intuition legitimate:

- What exactly *is* the area under a general curve $y = f(x)$? We can find the areas of simple shapes such as rectangles, triangles and circles without issue, but we haven't even defined what the area of a more complicated region should be.

- What is the precise relationship between antiderivatives and areas? Our intuition above suggests that antiderivatives should be related to areas, but as we saw in Theorem 5.6.3 a function can have infinitely many antiderivatives. Which one represents the area under the curve?

The first issue will be addressed in Section 6.2, where we use the concept of a Riemann sum to give a precise meaning to the notion of area under a curve, which we identify as a definite integral. The second issue will be addressed in Section 6.4 and forms the content of Fundamental Theorem of Calculus.

6.1 Areas Under Curves

We know how to find areas of simple shapes, such as rectangles, triangles and circles. But how can we find the area of a more complicated shape? Say, for example, the area under $y = x^3$ from $x = 0$ to $x = 1$? Well, if we don't know it exactly then we can at least approximate it using a shape we do know. For example, using rectangles.

A simple way to divide the graph of $y = x^3$ using rectangles is to pick several equally spaced points in the interval $[0, 1]$, say

$$x_0 = 1, x_1 = \frac{1}{3}, x_2 = \frac{2}{3}, \quad \text{and} \quad x_3 = 1,$$

and then draw rectangles using these points to determine the base of each rectangle and the height of the graph above the points to determine the height of each rectangle. In particular, we will have three rectangles; the first with base $[0, \frac{1}{3}]$, the second with base $[\frac{1}{3}, \frac{2}{3}]$ and the third with base $[\frac{2}{3}, 1]$. Now, the height of each rectangle comes from a point on the graph above the corresponding interval. To make these specific (and relatively straightforward), we'll use **right endpoints** to determine the height of each rectangle. So the height of each respective rectangle is given by $(\frac{1}{3})^3$, $(\frac{2}{3})^3$ and by 1^3.

Now, if we add up the areas of these rectangles we get an approximation for the area under the graph. Call these areas A_1, A_2, A_3. So the area under the graph of $y = x^3$ from $x = 0$ to $x = 1$ is about

$$A_1 + A_2 + A_3 = \frac{4}{9}.$$

There's nothing stopping us from going further here. For example, we pick n rectangles, each with base length $\frac{1}{n}$ and heights determined by right endpoints as above, then the area R_n of all of the rectangles will be an approximation (in fact, an overestimate) of the area of $y = x^3$. The bases of these rectangles are given by

$$[0, \frac{1}{n}], [\frac{1}{n}, \frac{2}{n}], \ldots, [\frac{n-1}{n}, \frac{n}{n}].$$

So the areas of these rectangles are

$$\frac{1}{n}\left(\frac{1}{n}\right)^3, \frac{1}{n}\left(\frac{2}{n}\right)^3, \ldots, \frac{1}{n}\left(\frac{n-1}{n}\right)^3, \frac{1}{n}\left(\frac{n}{n}\right)^3.$$

Adding these up, we get

$$R_n = \frac{1}{n}\frac{1^3 + 2^3 + 3^3 + \cdots + n^3}{n^3}$$

or just

$$R_n = \frac{1^3 + 2^3 + \cdots + n^3}{n^4}.$$

So if we sent n to ∞, we should get the exact area. In other words, the area under $y = x^3$ from $x = 0$ to $x = 1$ is given by

$$\lim_{n\to\infty} R_n = \lim_{n\to\infty} \frac{1^3 + 2^3 + \cdots + n^3}{n^4}.$$

Actually finding this limit requires some trickery. It's not at all obvious, but we can write

$$1^3 + 2^3 + \cdots + n^3 = \left[\frac{n(n+1)}{2}\right]^2.$$

Hence

$$\begin{aligned}
\lim_{n\to\infty} R_n &= \lim_{n\to\infty} \frac{1}{n^4}\left[\frac{n(n+1)}{2}\right]^2 \\
&= \lim_{n\to\infty} \frac{1}{4}\frac{n^2(n^2+2n+1)}{n^4} \\
&= \lim_{n\to\infty} \frac{1}{4}\frac{n^4+2n^3+n^2}{n^4} \\
&= \frac{1}{4}.
\end{aligned}$$

So the exact area under $y = x^3$ from $x = 0$ to $x = 1$ is $\frac{1}{4}$. Note that this matches up with our intuition from the last chapter: the area should be equal to the net change of an antiderivative. In this case, an antiderivative of x^3 is $\frac{1}{4}x^4$, and the net change of $\frac{1}{4}x^4$ from $x = 0$ to $x = 1$ is exactly $\frac{1}{4}$.

The process we used above for $y = x^3$ can be carried out for any continuous function. If we have a function defined on the interval $[a, b]$, we can approximate the area under the graph by using rectangles. First, we split $[a, b]$ into n different subintervals $[x_0, x_1], \ldots, [x_{n-1}, x_n]$, each of width $\Delta x = \frac{b-a}{n}$ and where

$$\begin{aligned}
x_0 &= a \\
x_1 &= a + \Delta x = a + \frac{b-a}{n} \\
x_2 &= a + 2\Delta x = a + 2\frac{b-a}{n} \\
&\vdots \\
x_n a &+ n\Delta x = b
\end{aligned}$$

The approximate area under the graph of $y = f(x)$ from $x = a$ to $x = b$ is then

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x.$$

We can now define the area under a graph of a continuous function. Again, the idea is to approximate the graph with more and more rectangles.

Definition 6.1.1 Area. The **area** under the graph of a continuous function $y = f(x)$ is the limit of the sum of the areas of the approximating rectangles:

$$A = \lim_{n\to\infty} R_n = \lim_{n\to\infty} [f(x_1)\Delta x + \cdots + f(x_n)\Delta x].$$

◊

Although the previous definition used right endpoints to find the height of each rectangle, this doesn't have to be the case. We can use left endpoints, midpoints or any other **sample points** $x_1^*, x_2^*, \ldots, x_n^*$, where each x_i^* is a point in the interval $[x_{i-1}, x_i]$.

Since finding areas in this way requires writing complicated sums, it seems like a good idea at this point to introduce notation for writing large summations. We call this **sigma notation**. Sigma notation takes the following form:

$$\sum_{i=\text{start}}^{\text{finish}} (\text{terms involving } i).$$

To evaluate such an expression, just plug in the different values for i from start to finish and add up what you get.

Example 6.1.2 Evaluating a sum. Determine $\sum_{i=-2}^{3}(4-3i)$.

Solution. We start at $i = -2$ and finish at $i = 3$. Plugging all of these values in for i, we get the following table:

Table 6.1.3 Table of values for $\sum_{i=-2}^{3}(4-3i)$

i	$4-3i$
-2	10
-1	7
0	4
1	1
2	-2
3	-5

So we can say that

$$\begin{aligned}\sum_{i=-2}^{3}(4-3i) &= 10+7+4+1-2-5\\ &= 15.\end{aligned}$$

□

Example 6.1.4 Using sigma notation. Use sigma notation to write the sum of all squares from 31 to 245.

Solution. We need to start at $i = 31$ and stop at $i = 245$. So the first component of this sum using sigma notation looks like $\sum_{i=31}^{245}$. Now we just need to figure out what goes inside of it. Since we're adding squares of numbers, we'll just put i^2 inside of this sum. Hence the sum of all squares from 31 to 245 can be denoted

$$\sum_{i=31}^{245} i^2.$$

□

Sigma notation has some useful properties, since it's just another way to write sums.

Theorem 6.1.5 Properties of Sigma Notation. *Let* $X = \sum_{i=a}^{b} x_i$ *and let* $Y = \sum_{i=a}^{b} y_i$. *Let* c *be a constant. Then the following are true:*

$$\sum_{i=a}^{b} c = c + \cdots + c = (b - a + 1)c$$

$$\sum_{i=a}^{b} cx_i = c\sum_{i=a}^{b} x_i = cX$$

$$\sum_{i=a}^{b}(x_i + y_i) = \sum_{i=a}^{b} x_i + \sum_{i=a}^{b} y_i = X + Y$$

$$\sum_{i=a}^{b}(x_i - y_i) = \sum_{i=a}^{b} x_i - \sum_{i=a}^{b} y_i = X - Y$$

$$\sum_{i=a}^{k} a_i + \sum_{i=k+1}^{b} a_i = \sum_{i=a}^{b} a_i.$$

In other words, we can break sigma notation up over addition and subtraction, and we can move constants outside of it.

Example 6.1.6 Areas using sigma notation. Write down a limit that gives the area under $f(x) = \sin x$ from $x = 1$ to $x = 3$ using sigma notation.

Solution. First, note that we can rewrite Definition 6.1.1 using sigma notation as

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x.$$

We know what $f(x)$ is, so we just need to identify x_i and Δx. Δx isn't too bad here, it's just

$$\Delta x = \frac{3-1}{n} = \frac{2}{n}.$$

The x_i are a little more complicated, but we'll still use right endpoints to find them. So we have

$$\begin{aligned} x_0 &= 1 \\ x_1 &= 1 + \frac{2}{n} = \frac{n+2}{n} \\ x_2 &= 1 + 2\frac{2}{n} = \frac{n+4}{n} \\ x_3 &= 1 + 3\frac{2}{n} = \frac{n+6}{n} \\ &\vdots \\ x_n &= 3 \end{aligned}$$

In general,

$$x_i = 1 + i\frac{2}{n} = \frac{n+2i}{n}$$

for $1 \le i \le n$.

Now we have everything we need to write our limit. Hence the area under $\sin x$ from 1 to 3 is given by

$$\lim_{n\to\infty} \sum_{i=1}^{n} \sin\left(\frac{n+2i}{n}\right)\frac{2}{n}.$$

□

Just as the problem of finding the slope of a tangent line can be viewed as a rate of change problem, the problem of finding areas can be viewed as an *accumulation problem*.

Example 6.1.7 Finding work done. In physics, the work done by a constant force F on some particle that moves a displacement d is given by the formula

$$W = Fd.$$

How should we define the work done by a variable force?

Solution. The main idea here is to treat the variable force as the combination of several constant forces. To be specific, let $F(x)$ denote our force (which depends on position x) and suppose it acts on a particle moving from $x = a$ to $x = b$. $F(x)$ itself may not be constant, but if we break $[a, b]$ up into n subintervals of the form $[x_{i-1}, x_i]$ each of length $\Delta x = \frac{b-a}{n}$, then $F(x)$ should be nearly constant on each subinterval. So we can pick points x_i^* from each subinterval and say that $F(x) \approx F(x_i^*)$ for all x in $[x_{i-1}, x_i]$.

Now we can approximate the work done by F as the particle moves through the subinterval $[x_{i-1}, x_i]$. The work done should be about equal to

$$F(x_i^*)\Delta x.$$

Now, if we add up these estimates for each subinterval, we see that the total work done should be approximated by

$$\sum_{i=1}^{n} F(x_i^*)\Delta x,$$

which suggests that we should define the exact work done to be

$$W = \lim_{n\to\infty} \sum_{i=1}^{n} F(x_i^*)\Delta x.$$

□

Note that Example 6.1.7 was not written in terms of finding areas, but still involved the limit from Definition 6.1.1. This tells us that the ideas in this section have a wider use than in just finding areas under curves.

6.2 The Definite Integral

The process used in Section 6.1 can be generalized quite a bit. Start with a function $f(x)$ defined on some interval $[a, b]$, and divide $[a, b]$ into n subintervals

$$[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]$$

where $x_0 = a, x_n = b$ and each subinterval has length Δx_k (i.e. $\Delta x_k = x_k - x_{k-1}$). This is known as a **partition** of $[a, b]$. Then we can choose sample points x_1^* from $[x_0, x_1]$, x_2^* from $[x_1, x_2]$, all the way up to x_n^* from $[x_{n-1}, x_n]$. Note that the only restriction we're placing on the sample point x_k^* is that is must come from the interval $[x_{k-1}, x_k]$, but it may be *any* point in this interval.

The partition and sample points selected are now used to determine a sum that represents an approximation to the area under the curve:

$$\sum_{i=1}^{n} f(x_k^*)\Delta x_k.$$

Geometrically, this sum represents the areas of multiple rectangles added together. Since these sums are essential for what follows, we will give them a name.

Definition 6.2.1 Riemann Sum. Let $a, b \in \mathbb{R}$ with $a < b$ and let $f(x)$ be a function defined on the interval $[a, b]$. Let $P = \{x_k\}_{k=0}^{n}$ be a partition of $[a, b]$ and choose corresponding sample points $x_k^* \in [x_{k-1}, x_k]$ for $1 \leq k \leq n$. Then a **Riemann sum** is a sum of the form

$$\sum_{k=1}^{n} f(x_k^*)\Delta x_k$$

where $\Delta x_k = x_k - x_{k-1}$. ♢

Although this can be more complicated than our definition in Definition 6.1.1, it still approximates the area under $f(x)$ from $x = a$ to $x = b$. And as each subinterval in this partition gets smaller (i.e. as $\Delta x_k \to 0$ or equivalently as $n \to \infty$), then this approximation should become exact, at least if $f(x)$ is "nice enough" (see Theorem 6.2.3 for a more precise statement).

Definition 6.2.2 The Definite Integral. Let $f(x)$ be a function defined on $[a, b]$. Then the **definite integral of $f(x)$ from a to b** is the number

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_k^*)\Delta x_k,$$

provided that the limit exists (and that $\Delta x_i \to 0$). If this limit exists, we say that $f(x)$ is **integrable** on $[a, b]$. ♢

We can use Definition 6.2.2 to rewrite our area definition: the area under $y = f(x)$ from $x = a$ to $x = b$ is defined to be the definite integral of f on $[a, b]$, assuming it exists. For a function to be integrable, the particular partition of $[a, b]$ that we use *cannot affect the limit* in Definition 6.2.2. This gives us a great deal of freedom in approximating the definite integral/area under a curve: when choosing sample points, we can use left endpoints; right endpoints; midpoints and more. However, if f is integrable then we lose nothing by just using right endpoints and subintervals of equal length, as in Definition 6.1.1.

Theorem 6.2.3 Integrability of Piecewise Continuous Functions. *Let $f(x)$ be a bounded, piecewise continuous function on $[a, b]$ with finitely many jump discontinuities. Then f is integrable on $[a, b]$. In particular, continuous functions are always integrable on bounded, closed intervals.*

All of the functions that we will work with will be piecewise continuous, and many important quantities in mathematics and its applications can be represented by bounded, piecewise continuous functions. Definite integrals of such quantities therefore exist and are often used as a measure of "accumulation." In particular, Example 6.1.7 can be generalized to provide the following definition of work done by a force.

Definition 6.2.4 Work. Let $F(x)$ denote the force acting on a particle over $[a, b]$. The **work** done by the force $F(x)$ on the particle as it moves from $x = a$ to $x = b$ is defined to be

$$W = \int_a^b F(x)\, dx\ .$$

♢

Example 6.2.5 Expressing work done. A particle moving from $x = -1$ to $x = 3$ is acted upon by the force $F(x) = 3x$. Find the work done by the force on this particle.

Solution. Using Definition 6.2.4, we see that the work done should be

$$W = \int_{-1}^{3} 3x\,dx.$$

Since the definite integral represents area, and the definite integral of $3x$ exists (since this is continuous), then we can see from Figure 6.2.6 that

$$W = \int_{-1}^{3} 3x\,dx = \frac{1}{2}(1)(-3) + \frac{1}{2}(3)(9) = 12.$$

Note that regions underneath the x-axis correspond to negative areas!

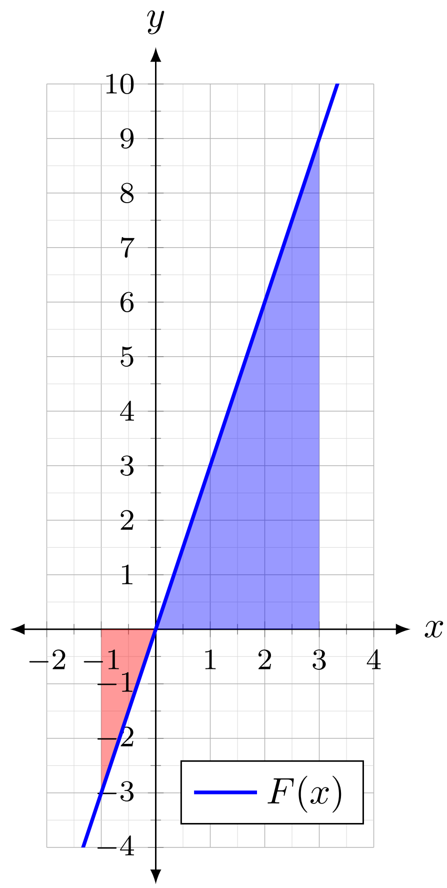


Figure 6.2.6 A graph of the force $F(x) = 3x$ and shaded areas corresponding to work.

□

In Example 6.2.5 we used the connection between definite integrals and areas to find the work done. For the moment, this is our only method for calculating definite integrals. We will develop other methods starting in Section 6.4, but the connection between definite integrals and areas is an important one that we will make use of frequently.

Example 6.2.7 A radical integral. Determine $\int_0^2 \sqrt{4-x^2}\,dx$.

Solution. We won't be able to solve this integral algebraically until Calculus 2, but we can solve it geometrically instead. If we graph $y = \sqrt{4-x^2}$, we get the upper half of the circle of radius 2 centered at the origin, as seen in

Figure 6.2.8. So the integral given must be exactly one-fourth the area of this circle, so

$$\int_0^2 \sqrt{4-x^2}\,dx = \frac{1}{4}\pi(2^2) = \pi.$$

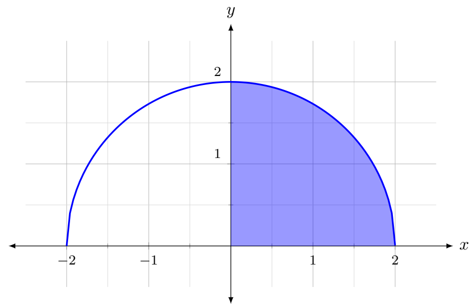


Figure 6.2.8 The graph of $y = \sqrt{4-x^2}$.

□

Since the definite integral is defined as a limit of sums, it shares many properties with sums. In particular, we have the following version of Theorem 6.1.5

Theorem 6.2.9 Properties of the Definite Integral. *Let $f(x)$ and $g(x)$ be integrable on $[a,b]$ and let c be a constant. Let t be any number in $[a,b]$. Then the following are true:*

$$\int_a^b c\,dx = c(b-a)$$
$$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$$
$$\int_a^b (f(x)+g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$
$$\int_a^b (f(x)-g(x))\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx$$
$$\int_a^t f(x)\,dx + \int_t^b f(x)\,dx = \int_a^b f(x)\,dx.$$

In other words, we can break definite integrals up over addition and subtraction, and we can move constants outside of it.

Computing integrals in practice often involves setting up and approximating Riemann sums. These sums depend on how we choose the sample points x_k^* from each interval $[x_{k-1}, x_k]$. We've mostly worked with $x_k^* = x_k$, or right endpoints. We could also choose $x_k^* = x_{k-1}$, using left endpoints. However, it's often more accurate to use *midpoints*: $x_k^* = \frac{x_{k-1}+x_k}{2}$. The reason for this is that if we choose midpoints of intervals as our sample points, these tend to be less affected than either endpoint if the function is increasing or decreasing. In particular, we have the **midpoint rule**.

Definition 6.2.10 Midpoint Rule. Let $f(x)$ be a continuous function. Then

$$\int_a^b f(x)\,dx \approx M_n = \sum_{k=1}^n f(x_k^*)\Delta x$$

where

$$x_k^* = \frac{x_{k-1}+x_k}{2}$$

$$\Delta x = \frac{b-a}{n}.$$

Note that this rule calls for evenly dividing $[a, b]$ into a partition $[x_{k-1}, x_k]$ where each subinterval has equal length Δx. ♢

Replacing the midpoints in Definition 6.2.10 with corresponding left endpoints (respectively, right endpoints) gives the **left endpoint rule** (respectively, the **right endpoint rule**). Together with the midpoint rule, these approximations are

$$L_n = \sum_{k=1}^{n} f(x_{k-1})\Delta x$$
$$M_n = \sum_{k=1}^{n} f\left(\frac{x_{k-1}+x_k}{2}\right)\Delta x$$
$$R_n = \sum_{k=1}^{n} f(x_k)\Delta x.$$

Careful! In general, $M_n \neq \frac{L_n+R_n}{2}$.

Example 6.2.11 Approximating an integral. Approximate $\int_1^3 \frac{1}{x}\,dx$ using Riemann sums with $n = 4$.

Solution. For this integral, we have

$$\Delta x = \frac{3-1}{4} = \frac{1}{2}$$
$$x_0 = 1$$
$$x_1 = \frac{3}{2}$$
$$x_2 = 2$$
$$x_3 = \frac{5}{2}$$
$$x_4 = 3$$

We will have different estimates using left endpoints, right endpoints and midpoints. Using left endpoints, we have

$$\int_1^3 \frac{1}{x}\,dx \approx L_4 = \frac{1}{2}\left(1 + \frac{2}{3} + \frac{1}{2} + \frac{2}{5}\right) = 1.283....$$

Using midpoints, we have

$$\int_1^3 \frac{1}{x}\,dx \approx M_4 = \frac{1}{2}\left(\frac{4}{5} + \frac{4}{7} + \frac{4}{9} + \frac{4}{11}\right) \approx 1.09.$$

Using right endpoints, we have

$$\int_1^3 \frac{1}{x}\,dx \approx R_4 = \frac{1}{2}\left(\frac{2}{3} + \frac{1}{2} + \frac{2}{5} + \frac{1}{3}\right) = 0.95.$$

Geometrically, we see that

$$0.95 = R_4 < \int_1^3 \frac{1}{x}\,dx < L_4 = 1.283...$$

The exact value of this integral is actually $\ln 3$, and we see that M_4, the midpoint estimate, is clearly the best approximation to the integral in this case. □

As n gets larger, ideally the approximations produced by the midpoint and endpoint rules become more exact. The figure below is adapted from an example by Marshall Hampton located at wiki.sagemath.org/interact/calculus and approximates integrals of the form $\int_a^b f(x)\,dx$. This makes use of the computer algebra system Sage which is itself based on the Python language.

Specify a static image with the `@preview` attribute;
Or create and provide an automatic screenshot as
`generated/preview/interactive-numerical-integral-preview.png`
via the `PreTeXt-CLI` application or `pretext/pretext` script.

Figure 6.2.12 Numerical integrals using the midpoint, left endpoint and right endpoint rules

The geometric interpretation of definite integrals as areas also suggests the following formulas. These can be useful for simplifying and estimating certain definite integrals.

Theorem 6.2.13 More Properties of the Definite Integral. *Let $f(x), g(x)$ be integrable functions on $[a, b]$. Then*

1. $\displaystyle\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$
2. $\displaystyle\int_a^a f(x)\,dx = 0$
3. *If $f(x) \geq g(x)$ on $[a, b]$, then $\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$.*
4. *If $m \leq f(x) \leq M$ on $[a, b]$, then $m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a)$.*

6.3 Evaluating Integrals

Just as we did with derivatives, we want to find a way to calculate integrals without having to compute limits every time.

Theorem 6.3.1 Evaluation Theorem. *Let $f(x)$ be continuous on $[a, b]$ and let $F(x)$ be a single antiderivative of $f(x)$ on this interval. Then*

$$\int_a^b f(x)\,dx = F(a) - F(b).$$

Proof. First, partition $[a, b]$ into $[x_0, x_1], \ldots, [x_{n-1}, x_n]$. Then we can say that

$$\begin{aligned} F(b) - F(a) &= F(x_n) - F(x_0) \\ &= F(x_n) - F(x_{n-1}) + F(x_{n-1}) - F(x_{n-2}) + \cdots + F(x_1) - F(x_0) \\ &= \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})]. \end{aligned}$$

By the Mean Value Theorem, we can rewrite each term in the sum using F':

$$F(x_i) - F(x_{i-1}) = F'(x_i^*)(x_i - x_{i-1})$$

where x_i^* is just some point in $[x_{i-1}, x_i]$. So we have

$$\begin{aligned} F(b) - F(a) &= \sum_{i=1}^{n} F'(x_i^*)(x_i - x_{i-1}) \\ &= \sum_{i=1}^{n} f(x_i^*)\Delta x. \end{aligned}$$

If we now send $n \to \infty$ (i.e„ $\Delta x \to 0$), then the above becomes

$$F(b) - F(a) = \int_a^b f(x)\,dx.$$

■

Theorem 6.3.1 tells us that if we want to evaluate the definite integral of a continuous function, we need to first find an antiderivative. We can also think of this result as follows: the definite integral of a rate of change is just the net change of the antiderivative.

Example 6.3.2 Modeling a population. A honeybee population starts with 100 members and grows at a rate of $n'(t)$ per week. Find the total population after 15 weeks in terms of $n'(t)$.

Solution. The total population after $15 weeks$ should be given by

$$\begin{aligned} \text{(total population)} &= \text{(initial population)} + \text{(net change)} \\ &= 100 + \int_0^{15} n'(t)\,dt \end{aligned}$$

□

As a useful shorthand, we introduce the notation $F(b) - F(a) = F(x)\big]_a^b$. So if F is an antiderivative of f, we can say that

$$\int_a^b f(x)\,dx = F(x)\big]_a^b.$$

Example 6.3.3 Area under a rational function. Find the area under $f(t) = \frac{t^2-1}{t^4-1}\,dt$ from $t = 0$ to $t = \frac{1}{\sqrt{3}}$.

Solution. The area is just the definite integral, which is

$$\begin{aligned} \int_0^{\frac{1}{\sqrt{3}}} f(t)\,dt &= \int_0^{\frac{1}{\sqrt{3}}} \frac{t^2-1}{t^4-1}\,dt \\ &= \int_0^{\frac{1}{\sqrt{3}}} \frac{1}{t^2+1}\,dt \\ &= \tan^{-1} t\big]_0^{\frac{1}{\sqrt{3}}} \\ &= \tan^{-1}\left(\frac{1}{\sqrt{3}}) - \tan^{-1}(0)\right) \\ &= \frac{\pi}{6}. \end{aligned}$$

□

The **average value** of a continuous function $f(x)$ on an interval $[a, b]$ is defined to be

$$\frac{1}{b-a}\int_a^b f(x)\,dx.$$

Example 6.3.4 Average force. An object of mass 3 kg is acted upon by several forces that combine to give it an acceleration of $a(t) = 4e^{-t} + \sin t$, where $a(t)$ is in units of $\frac{\text{m}}{\text{s}^2}$ and t is in seconds. Find the average force acting on the mass from $t = 1$ to $t = 3$.

Solution. We first need to find the net force acting on the particle. By Newton's Second Law, this is $F_{\text{net}} = ma = 12e^{-t} + 4\sin t$. So the average force should be

$$\begin{aligned}\frac{1}{2}\int_1^3 (12e^{-t} + 4\sin t)\,dt &= \frac{1}{2}\big[-12e^{-t} - 4\cos t\big]_1^3 \\ &= \frac{-12e^{-3} - 4\cos 3 + 12e^{-1} + 4\cos 1}{2}.\end{aligned}$$

□

Because of the relationship between definite integrals and antiderivatives, we introduce a notation for antiderivatives that reinforces this connect.

Definition 6.3.5 Indefinite Integral. Let $f(x)$ be a continuous function. The **indefinite integral** of $f(x)$, denoted $\int f(x)\,dx$, is the general antiderivative of $f(x)$. In symbols, if $F(x)$ is an antiderivative of $f(x)$ then

$$\int f(x)\,dx = F(x) + C.$$

◊

Example 6.3.6 Finding an indefinite integral. Given that

$$\frac{d}{dx}[xe^x - e^x] = xe^x,$$

find $\int [xe^x + \sec^2 x]\,dx$.

Solution. We can find this integral by splitting it up, since the antiderivative of a sum is just the sum of the antiderivatives. So

$$\begin{aligned}\int [xe^x + \sec^2 x]\,dx &= \int xe^x\,dx + \int \sec^2 x\,dx \\ &= xe^x - e^x + \tan x + C.\end{aligned}$$

□

Note that using Theorem 6.3.1 to find definite integrals relies on having a particular antiderivative of a function $f(x)$, but we've said nothing about whether or not such an antiderivative can exist. In the next section, we'll see how to use integrals to construct antiderivatives.

6.4 The Fundamental Theorem of Calculus

The previous section used antiderivatives to find definite integrals. Now we'll go the other direction and use integrals to define antiderivatives. We can do this as follows.

Let f be some function continuous on $[a, b]$. Then we can say that f must be continuous on every interval of the form $[a, x]$ for x in $[a, b]$. This means that the definite integrals $\int_a^x f(t)\,dt$ exist, and so we can use these to define a function, say $g(x)$.

Example 6.4.1 Functions from integrals. Let $f(x) = 2 - x^2$ and define $g(x) = \int_1^x f(t)\,dt$. Find $g(x)$. Where is $g(x)$ the largest?

Solution. First, note that we can compute $\int_1^x f(t)\,dt$ by finding an antiderivative of $f(t) = 2 - t^2$. So

$$g(x) = \int_1^x (2 - t^2)\,dt = \left[2t - \frac{1}{3}t^3\right]_{t=1}^{x}$$

or just

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

We can find where g is largest by using a derivative test or just geometry. Since g represents the area under f from 1 to x, it follows that this area will be largest at $x = \sqrt{2}$, since this is where f crosses the axis. □

There's one important thing to take away from the previous example. The function $g(x) = \int_1^x f(t)\,dt$ turned out to be an antiderivative of $f(x)$, so in particular $g'(x) = f(x)$. This is not a coincidence.

Theorem 6.4.2 Fundamental Theorem of Calculus. *Let $f(x)$ be a continuous function defined on $[a, b]$.*

1. *Define $g(x) = \int_a^x f(t)\,dt$ for x in $[a, x]$. Then $g'(x) = f(x)$, i.e., g is an antiderivative of f.*

2. *Let F be a particular antiderivative of f. Then $\int_a^b f(x)\,dx = F(b) - F(a)$.*

Proof. We'll sketch the proof of the first part. First, assume that $g(x) = \int_a^x f(t)\,dt$. Then

$$\frac{g(x+h) - g(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,dt.$$

Now, if h is small and if $x \leq t \leq x + h$, then $\int_x^{x+h} f(t)\,dt$ should be approximately equal to $f(x)h$ since f is continuous. Therefore

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

and so $g'(x) = f(x)$. ■

The second part of Theorem 6.4.2 is something we've seen before in Theorem 6.3.1, but the first part is new. Essentially, this result says that integration and differentiation are *inverses* of each other.

Example 6.4.3 Error function. The **error function**, defined by

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt,$$

is one of the most important functions in statistics and other applications of mathematics (such as modeling the spread of a virus or the spread of heat). Find the derivative of $\mathrm{erf}(x)$.

Solution. By the Fundamental Theorem of Calculus, we have

$$erf'(x) = \frac{2}{\sqrt{\pi}}e^{-x^2}.$$

Note that $\mathrm{erf}'(x) \to 0$ as $x \to \infty$, which suggests that the error function "levels out" as x increases, which is in fact true. □

Example 6.4.4 Computing derivatives. Let $f(x) = \int_{\sin x}^{1} \sqrt{t^2 + e^{-t}}\, dt$. Find $f'(x)$.

Solution. We have

$$\begin{aligned} f'(x) &= \frac{d}{dx} \int_{\sin x}^{1} \sqrt{t^2 + e^{-t}}\, dt \\ &= -\frac{d}{dx} \int_{1}^{\sin x} \sqrt{t^2 + e^{-t}}\, dt \\ &= -\sqrt{\sin^2 x + e^{-\sin x}} \cos x. \end{aligned}$$

Here, $\sin x$ acts like an inside function so we must use the chain rule in addition to the Fundamental Theorem of Calculus to find this derivative. □

Example 6.4.5 A business problem. A company purchases equipment that depreciates at a rate $f(t)$ and accumulates maintenance costs at a rate $g(t)$, where t denotes the number of months after purchase. The company wants to minimize the cost per month of the equipment due to depreciation and maintenance. After how much time should the company replace the equipment?

Solution. First, we need to figure out the cost per month due to depreciation and maintenance, call it $C(t)$. Then

$$C(t) = \frac{\text{total depreciation} - \text{total maintenance}}{t}.$$

Now, the total depreciation after t months is $\int_0^t f(s)\, ds$, and the total maintenance cost after t months is $\int_0^t g(s)\, ds$. So

$$C(t) = \frac{1}{t} \int_0^t [f(s) + g(s)]\, ds.$$

This is the function we need to minimize.

To minimize $C(t)$, we'll start by finding its critical points. Assuming that each of these functions is differentiable, we have

$$C'(t) = -\frac{1}{t^2} \int_0^t [f(s) + g(s)]\, ds + \frac{f(t) + g(t)}{t}$$

which reduces to

$$C'(t) = -\frac{1}{t} C(t) + \frac{f(t) + g(t)}{t}.$$

So the critical points occur at any value $t = T$ for which $C(T) = f(T) + g(T)$, i.e., the cost per month due to depreciation and maintenance equals the rate of depreciation plus the rate of maintenance costs.

Now we need to see if we have a minimum or a maximum. To do so, we can use the second derivative test. After some algebra, we get that

$$C''(t) = 2\frac{C(t) - (f(t) + g(t))}{t^2} + \frac{f'(t) + g'(t)}{t^2},$$

and so

$$C''(T) = \frac{f'(T) + g'(T)}{T^2} > 0.$$

Therefore T gives a local minimum. If we assume that $C(t)$ is differentiable everywhere, then this local minimum should also be an absolute minimum since it's the only critical point. □

6.5 Substitution

By Theorem 6.4.2, we know that integration and differentiation are inverse operations. This means that differentiation rules, such as the chain rule and product rule, have corresponding *integration* rules. In this section we look at the integration rule corresponding to the chain rule, known as u-substitution, and in Calculus 2 you'll see the integration rule corresponding to the product rule.

To see how the chain rule leads to an integration rule, suppose we have an expression $F(g(x))$ where F, g are differentiable. Then the chain rule says that $\frac{d}{dx}F(g(x)) = F'(g(x))g'(x)$. Now, write $F' = f$ and make the substitution $u = g(x)$. Then we get

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

Now we can integrate both sides of this equation with respect to x to get

$$\int f(g(x))g'(x)\,dx = \int f(u)\frac{du}{dx}\,dx = \int f(u)\,du.$$

This method is known as **u-substitution**.

Theorem 6.5.1 u-Substitution. *Let $u = g(x)$ be a differentiable function whose range is an interval I. Let $f(x)$ be a continuous function on I. Then*

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du.$$

Solving an integral $\int f(x)\,dx$ by u-substitution typically proceeds as follows:

1. Choose some substitution u to simplify your integral. Your choice for u will most likely be an "inside function".
2. Rewrite your integral in terms of u and du instead of x and dx.
3. Solve the resulting integral in terms of u if possible, and then rewrite your answer in terms of the original variable x.

Example 6.5.2 A simple substitution. Determine $\int e^{3x-2}\,dx$ by using the substitution $u = 3x - 2$.

Solution. We need to rewrite the integral in terms of u. An easy way to start is to replace $3x - 2$ with u:

$$\int e^{3x-2}\,dx = \int e^u\,dx.$$

However, *we can't integrate yet.* This is because the integral needs to be written entirely in terms of u and du; no terms involving x, including dx, can remain. to rewrite dx in terms of du, we'll find $\frac{du}{dx}$ and then "solve" for dx:

$$\frac{du}{dx} = 3 \Rightarrow du = 3\,dx \Rightarrow dx = \frac{du}{3}.$$

So our integral turns into

$$\begin{aligned}\int e^{3x-2}\,dx &= \int e^u\left(\frac{du}{3}\right)\\ &= \frac{1}{3}\int e^u\,du\end{aligned}$$

$$= \frac{1}{3}e^u + C.$$

The last step is to replace u with $3x-2$, and so

$$\int e^{3x-2}\,dx = \frac{1}{3}e^{3x-2} + C.$$

We can also check this by differentiating. □

Example 6.5.3 Simplifying an integral with substitution. Determine $\int (2-5t)^{-\frac{3}{2}}\,dt$.

Solution. We can find $\int u^{-\frac{3}{2}}\,du$ without too much trouble, but this integral is more complicated. However, we can use u-substitution to get our integral into this form. So we'll let $u = 2-5t$. Then $du = -5\,dt$, which means $dt = -\frac{du}{5}$ and so

$$\begin{aligned}\int (2-5t)^{-\frac{3}{2}}\,dt &= \int u^{-\frac{3}{2}}\left(-\frac{1}{5}\,du\right)\\ &= -\frac{1}{5}\int u^{-\frac{3}{2}}\,du\\ &= \frac{1}{5}\frac{2}{1}u^{-\frac{1}{2}} + C\\ &= \frac{2}{5}(2-5t)^{-\frac{1}{2}} + C.\end{aligned}$$

□

u-substitution is often useful if one part of an integral can be written as the derivative of another part of the integral.

Example 6.5.4 Integral of tangent. Find the general antiderivative of $\tan x$.

Solution. We need to find $\int \tan x\,dx$. If we rewrite $\tan x$ in terms of sine and cosine, we get

$$\int \tan x = \int \frac{\sin x}{\cos x}\,dx.$$

It's not obvious, but we can solve this using u-substitution since $\sin x$ looks a lot like the derivative of $\cos x$. So we'll let $u = \cos x$, giving $du = -\sin x\,dx$ or just $dx = -\frac{du}{\sin x}$. This simplifies our integral to

$$\int \tan x\,dx = \int -\frac{du}{u} = -\ln|u| + C = -\ln|\cos x| + C.$$

Since $-\ln|\cos x| = \ln|\cos x|^{-1} = \ln|\sec x|$, this shows that the general antiderivative of $\tan x$ is

$$\int \tan x\,dx = \ln|\sec x| + C.$$

Note that $u = \sin x$ *would not work* in this example, at least not without some nasty algebra. □

We can also use u-substitution to find definite integrals.

Example 6.5.5 Integrating with a natural log. Determine $\int_{e^2}^{e^4} \frac{dx}{x\sqrt{\ln x}}$.

Solution. This integral looks awful, but the integrand contains $\ln x$ and its derivative $\frac{1}{x}$, which suggests that the substitution $u = \ln x$ might simplify things. So let $u = \ln x$. Then $du = \frac{1}{x}\,dx$, or $dx = x\,du$, and the integral becomes

$$\begin{aligned}
\int_{e^2}^{e^4} \frac{dx}{x\sqrt{\ln x}} \int_{x=e^2}^{x=e^4} \frac{x\,du}{x\sqrt{u}} \\
&= \int_{x=e^2}^{x=e^4} \frac{du}{\sqrt{u}} \\
&= \Big[u\Big]_{x=e^2}^{x=e^4} \\
&= \Big[\ln x\Big]_{e^2}^{e^4} \\
&= 2.
\end{aligned}$$

□

In the last example we had to be careful about our limits of integration since they applied to x but not u. We can simplify the process a little by converting the x limits to u limits, as shown in the next example.

Example 6.5.6 Area under $x \sin x^2$. Find the area under the graph of $y = x \sin x^2$ from $x = 0$ to $x = \sqrt{\pi}$.

Solution. The area is given by $\int_0^{\sqrt{\pi}} x \sin x^2 \, dx$. We can solve this by u-substitution. Let $u = x^2$, our inside function. Then $du = 2x\,dx$, so $dx = \frac{du}{2x}$ and our integral turns into

$$\int_{x=0}^{x=\sqrt{\pi}} \frac{1}{2} \sin u \, du.$$

Now we change the limits to be in terms of u, giving us $u = 0$ and $u = \pi$. So the area is just

$$\begin{aligned}
\int_0^{\sqrt{\pi}} x \sin x^2 \, dx &= \int_0^{\pi} \frac{1}{2} \sin u \, du \\
&= \frac{1}{2}[-\cos u]_0^{\pi} \\
&= \frac{1}{2}(\cos 0 - \cos \pi) \\
&= 1.
\end{aligned}$$

□

Chapter 7

Techniques of Integration

This chapter focuses on several techniques for evaluating complicated integrals. As the tools in this chapter are in general more complicated than those that appeared in Chapter 6, you should only use these if algebraic simplification or substitution are insufficient to compute an integral.

7.1 Integration by Parts

Suppose we want to integrate $f(x) = x \sin x$. This can't be simplified using algebra (since integrals don't split up over products!) and substitution probably won't be very helpful, since $f(x)$ does not already contain an "inside function" that we could replace. The integration tools that we have are unable to compute this integral, so we need to develop more tools!

Recall that the substitution method for solving integrals from Section 6.5 is derived from the chain rule. In fact, by Theorem 6.4.2 every derivative rule available to us also provides a corresponding integration rule. So perhaps the integration rule corresponding to the product rule can help us to integrate products. We call this method *integration by parts.*

Theorem 7.1.1 Integration by Parts. *Let $f(x)$ and $g(x)$ be continuous functions, with $g(x)$ differentiable. Then*

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int g(x)f'(x)\,dx.$$

Proof. By the product rule, we have $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + f'(x)g(x)$. Rearranging this, we have

$$f(x)g'(x) = [f(x)g(x)]' - f'(x)g(x).$$

Finally, we can integrate both sides to get

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int g(x)f'(x)\,dx.$$

■

A useful way to remember the integration by parts method is as follows:

$$\int u\,dv = uv - \int v\,du. \tag{II.1}$$

Integration by parts can be used to integrate products of functions *that can't already be evaluated by simpler means.*

Example 7.1.2 Using the Integration by Parts Formula. Compute $\int x \sin x \, dx$.

Solution. To use (II.1), we need to write $\int x \sin x \, dx$ in the form $\int u \, dv$. Once we have u and dv, then we can set up the rest of the formula. One thing that can make this method tricky is the fact that we can have many different choices for u and dv; the only thing we need for certain is for $u \, dv = x \sin x$ here. However, if we look at the formula (II.1) then we'll see two new terms on the right hand side: du and v, together in the second integral. This means that we'll need to differentiate u and integrate v, and this will suggest what both should be.

Since we need to compute du, then we want u to be easy to differentiate. Similarly, we want dv to be easy to integrate. And since we'd like to be able to compute this integral, we'd also like $v \, du$ to be simpler than $u \, dv$. Now return to $x \sin x$. Since x gets simpler if we differentiate it by $\sin x$ doesn't, this suggests that we should set $u = x$. And since $u = x$, we must have $dv = \sin x \, dx$ (since dv automatically becomes whatever is left over). With these choices, we have

$$\begin{aligned} du &= dx \\ v &= -\cos x \end{aligned}$$

Therefore

$$\begin{aligned} \int x \sin x \, dx &= \int u \, dv \\ &= uv - \int v \, du \\ &= -x \cos x - \int (-\cos x) \, dx \\ &= -x \cos x + \int \cos x \, dx \\ &= -x \cos x + \sin x + C. \end{aligned}$$

□

You may be wondering why we didn't include an arbitrary constant when we found v in Example 7.1.2. This is because including an arbitrary constant here doesn't affect the final answer. It's certainly not incorrect to include one here, but it adds more work than is already necessary.

When using the integration by parts method, it's usually a good idea to choose u to be the polynomial portion of your integrand if there is one. This is because polynomials become simpler when they're differentiated. However, there are examples where u should be something else.

Example 7.1.3 Integrals Involving the Natural Logarithm. Determine $\int 3x^2 \ln x \, dx$.

Solution. Since we need to integrate a product, and we can't easily do so with u-substitution or algebra, we'll try integration by parts. We could pick $u = 3x^2$, and this is certainly easy to differentiate so we can find du. However, doing so leaves us with $dv = \ln x$, which is tricky to integrate, but is easy to differentiate. So we'll try setting $u = \ln x$ and $dv = 3x^2 \, dx$. Then

$$du = \frac{1}{x} \, dx$$

$$v = x^3$$

and so

$$\begin{aligned}\int 3x^2 \ln x\,dx &= uv - \int v\,du \\ &= x^3 \ln x - \int \frac{x^3}{x}\,dx \\ &= x^3 \ln x - \frac{1}{3}x^3 + C.\end{aligned}$$

□

Example 7.1.4 An Integral from Differential Equations. An important concept in the field of differential equations is that of the *Laplace transform*. In the process of computing the Laplace transform of $\cos t$, an integral similar to $\int e^t \cos t\,dt$ appears. Determine this integral.

Solution. Here, it doesn't make much difference what we select for u or dv, since both e^t and $\cos t$ are easy to differentiate and integrate, while neither becomes simpler. But since we're integrating a product, we'll try to use integration by parts anyway and hope for the best. To that end, let $u = e^t$ and $dv = \cos t$. Then $du = e^t\,dt$ and $v = \sin t$, which gives

$$\begin{aligned}\int e^t \cos t\,dt &= uv - \int v\,du \\ &= e^t \sin t - \int e^t \sin t\,dt.\end{aligned}$$

As expected, the new integral isn't any simpler than the one that we started with. However, it's not any more *complicated* either. So we'll try integration by parts on the new integral as well, with $u = e^t, du = e^t\,dt, dv = \sin t\,dt$ and $v = -\cos t$. Then

$$\int e^t \sin t\,dt = -e^t \cos t + \int e^t \cos t\,dt.$$

So we have

$$\begin{aligned}\int e^t \cos t\,dt &= e^t \sin t - (-e^t \cos t + \int e^t \cos t\,dt) \\ &= e^t \sin t + e^t \cos t - \int e^t \cos t\,dt\end{aligned}$$

There is something very strange about this equation: *the same integral appears on both sides*. This is no longer a calculus problem, it's an algebra problem! Solving for the integral algebraically (and including the arbitrary constant!) gives

$$\int e^t \cos t\,dt = \frac{e^t \sin t + e^t \cos t}{2} + C.$$

□

Integrals with integrands of the form $P(x)e^x$, $P(x)\sin x$ or $P(x)\cos x$ (where $P(x)$ is a polynomials) can be successfully evaluated using the *tabular method*.

Example 7.1.5 The Tabular Method. Evaluate $\int(3x^2 - 2x)\cos 3x\,dx$. □

Integration by parts also applies to definite integrals. You only need to remember to carry the limits through.

Example 7.1.6 Sign of a Sine. Let $f(x) = x^2 \sin x$. Does $f(x)$ tend to be positive or negative along $[0, 2\pi]$?

Solution. We need to check the sign of $\int_0^{2\pi} x^2 \sin x \, dx$. Using the tabular method, we get

$$\int_0^{2\pi} x^2 \sin x \, dx = [-x^2 \cos x + 2x \sin x + 2 \cos x]_0^{2\pi} = -4\pi^2 + 4\pi - 2.$$

Since this is negative, $f(x)$ tends to be negative along $[0, 2\pi]$. □

7.2 Trigonometric Integrals and Substitution

In this section we determine how to evaluate integrals involving powers of trigonometric functions (such as sin, sec and others), as well as integrals involving certain radicals. Much of our work will follow from the important identity $\sin^2 x + \cos^2 x = 1$, as well as the *half-angle formulas*

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

Integrals Involving Powers of Sine and Cosine

Example 7.2.1 Integrating an Odd Power of Cosine. Determine $\int_0^{\frac{\pi}{6}} \cos^5(3x) \, dx$.

Solution. Although it's not obvious, we can actually solve this integral with substitution. First, we'll separate a power of $\cos 3x$ from the rest in order to act as our du:

$$\int_0^{\frac{\pi}{6}} \cos^4(3x) \cos 3x \, dx.$$

Now, if we want $\cos 3x$ to act as du in a substitution, then we need u to look like $\sin 3x$. We can do this by using the Pythagorean Identity and writing $\cos^4(3x) = (1 - \sin^2(3x))^2$.

Now we'll set $u = \sin 3x$, which gives $du = 3 \cos 3x \, dx$ and

$$\begin{aligned} \int_0^{\frac{\pi}{6}} \cos^4(3x) \cos 3x \, dx &= \int_0^{\frac{\pi}{6}} (1 - \sin^2(3x))^2 \cos 3x \, dx \\ &= \frac{1}{3} \int_0^{\frac{1}{2}} (1 - u^2)^2 \, du \\ &= \int_0^{\frac{1}{2}} (1 - 2u^2 + u^4) \, du, \end{aligned}$$

which we can finish integrating without much trouble. □

The trick we used above only worked because we had an odd power of cosine. In general, when integrating $\int \sin^m x \cos^n x \, dx$, you want to separate a factor from an odd power (if there is one). Then finish by using substitution.

Example 7.2.2 Odd Powers of Sine and Cosine. Determine $\int \sin^3 t \cos^5 t \, dt$.

Solution. Remember: we want to separate a factor off of an odd power if there is one. Thankfully, we have two! We'll pick sine this time:

$$\begin{aligned}\int \sin^3 t \cos^5 t \, dt &= \int \sin^2 t \cos^5 t \sin t \, dt \\ &= \int (1-\cos^2 t) \cos^5 t \sin t \, dt \\ &= -\int (1-u^2) u^5 \, du \\ &= -\int (u^5 - u^7) \, du.\end{aligned}$$

At this point, the integral isn't too difficult to complete. □

If integrating even powers of sine and cosine, the algebra gets a little worse since the substitution trick applied above no longer works.

Example 7.2.3 Integrating Even Powers of Cosine. The integral $\int_0^{\frac{\pi}{2}} \cos^2\theta \, d\theta$ appears when finding the area of the circle using integration. Find the value of this integral.

Solution. Since we have an even power to work with, we'll try the half-angle formula. This lets us reduce the power on the cosine, at the expense of multiplying θ by 2, which greatly works in our favor:

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \cos^2\theta \, d\theta &= \int_0^1 \frac{1+\cos 2\theta}{2} \, d\theta \\ &= \frac{1}{2}\int_0^{\frac{\pi}{2}} (1+\cos 2\theta) \, d\theta \\ &= \frac{1}{2}\left[\theta + \frac{1}{2}\sin 2\theta\right]_0^{\frac{\pi}{2}} \\ &= \frac{\pi}{4}.\end{aligned}$$

□

Integrals Involving Powers of Secant and Tangent

In contrast to those involving sine and cosine, all of the integrals here involving secant and tangent are solvable using substitution. Generally, we'll attack these problems by separating a factor of $\sec^2 x$ (in which case we use $u = \tan x$) or a factor of $\sec x \tan x$ (in which case we use $u = \sec x$). This will often be done using the *Pythagorean identity*

$$\tan^2 x = \sec^2 x - 1$$

We will occasionally need to make use of two integral formulas:

$$\begin{aligned}\int \sec x \, dx &= \ln|\sec x + \tan x| + C \\ \int \tan x \, dx &= \ln|\sec x| + C\end{aligned}$$

Example 7.2.4 Integrating a Power of Tangent. Determine $\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^6 w \, dw$.

Solution. Our goal is to separate a factor of either $\sec^2 w$ or $\sec w \tan w$ to serve as our du. Since neither appears, we'll use the Pythagorean identity to introduce a $\sec^2 w$ into our integral:

$$\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^6 w\, dw = \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^4 w \sec^2 w\, dw - \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^4 w\, dw.$$

At this point the first integral can be solved using substitution, but the second integral needs to be rewritten again. However, the same trick we applied before works here too: just use the Pythagorean identity to introduce a factor of $\sec^2 w$. This gives

$$\begin{aligned}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^6 w\, dw &= \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^4 w \sec^2 w\, dw - \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^2 w \sec^2 w\, dw + \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^2 w\, dw.\\ &= \int_{-1}^{1} u^4\, du - \int_{-1}^{1} u^2\, du + \int_{-1}^{1} du - \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} dw\\ &= \frac{2}{5} - \frac{2}{3} + 2 - \frac{\pi}{2}.\end{aligned}$$

So

$$\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \tan^6 w\, dw = \frac{2}{5} - \frac{2}{3} + 2 - \frac{\pi}{2}.$$

□

Example 7.2.5 Integrating a Product of Tangent and Secant. Determine $\int \tan x \sec^3 x\, dx$

Solution. Here, we see that we can separate a factor of $\tan x \sec x$ from the integrand, so we'll do that and use $u = \sec x$:

$$\begin{aligned}\int \tan x \sec^3 x\, dx &= \int \sec^2 x \sec x \tan x\, dx\\ &= \int u^2\, du\\ &= \frac{\sec^3 x}{3} + C.\end{aligned}$$

□

In general, try to follow the guidelines below when integrating products of secant and tangent:

- If the power of secant is *even*, then separate a factor of $\sec^2 x$ and use $u = \tan x$, along with the Pythagorean identity if necessary.
- If the power of tangent is *odd*, then separate a factor of $\sec x \tan x$ and use $u = \sec x$, along with the Pythagorean identity if necessary.

Now, for an example that does not fall into the above guidelines.

Example 7.2.6 Integrating Secant Cubed. Determine $\int \sec^3 x\, dx$.

Solution. We're dealing with an odd power of secant, so splitting off a factor of $\sec^2 x$ won't help here. We also lack a $\tan x$ term, so splitting off a factor of $\sec x \tan x$ is tricky as well. We'll try integrating by parts instead, with

$$\begin{aligned}u &= \sec x & du &= \sec x \tan x\, dx\\ dv &= \sec^2 x\, dx & v &= \tan x\end{aligned}$$

This gives

$$\begin{aligned}
\int \sec^3 x\,dx &= \sec x \tan x - \int \sec x \tan^2 x\,dx \\
&= \sec x \tan x - \int \sec^3 x\,dx + \int \sec x\,dx \\
&= \sec x \tan x - \int \sec^3 x\,dx + \ln|\sec x + \tan x|
\end{aligned}$$

Therefore

$$\int \sec^3 x\,dx = \frac{1}{2}(\sec x \tan x + \ln|\sec x + \tan x|) + C.$$

□

Trigonometric Substitutions

An ellipse with horizontal axis length a and vertical axis length b centered at the origin is given by

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

If we want to find the area of this ellipse, then we can compute the following integral and multiply it by 4:

$$\int_0^a \sqrt{b^2 - (\frac{b}{a})^2 x^2}\,dx = \int_0^a \frac{b}{a}\sqrt{a^2 - x^2}\,dx.$$

This integral is difficult to solve due to the square root in the integrand; it's tough to imagine what an antiderivative might be. So what we'll try to do is to use a "reverse substitution" that simplifies the square root.

To do so, notice that the Pythagorean identity gives $\cos^2\theta = 1 - \sin^2\theta$. Taking the square root of the left hand side is much easier here than taking a root of the right hand side. So we'll try that substitution here: let

$$\begin{aligned}
x &= a\sin\theta \\
dx &= a\cos\theta\,d\theta
\end{aligned}$$

Then

$$\begin{aligned}
\int_0^a \sqrt{a^2 - x^2}\,dx &= \int_{x=0}^{x=a} \sqrt{a^2 - a^2\sin^2\theta}\,a\cos\theta\,d\theta \\
&= a^2 \int_{x=0}^{x=a} \cos^2\theta\,d\theta
\end{aligned}$$

Before we go further, let's figure out what our new limits are. Since $x = a\sin\theta$, we have $\theta = \sin^{-1}\frac{x}{a}$, and so $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Hence we need our limits for θ to be $\theta = 0$ and $\theta = \frac{\pi}{2}$, and so we get

$$\int_0^a \sqrt{a^2 - x^2}\,dx = a^2 \int_0^{\frac{\pi}{2}} \cos^2\theta\,d\theta.$$

We've seen how to integrate a function like this in Example 7.2.3. Hence the area of this ellipse is

$$4\int_0^a \frac{b}{a}\sqrt{a^2 - x^2}\,dx = 4ab\int_0^{\frac{\pi}{2}} \cos^2\theta\,d\theta = ab\pi.$$

This method is known as **trigonometric substitution**, and is useful for simplifying radicals. Table 7.2.7 shows the various substitutions that can be used.

Table 7.2.7 Trigonometric substitutions

radical	substitution	identity
$\sqrt{a^2 - x^2}$	$x = a\sin\theta$	$\cos^2\theta = 1 - \sin^2\theta$
$\sqrt{x^2 - a^2}$	$x = a\sec\theta$	$\tan^2\theta = \sec^2\theta - 1$
$\sqrt{a^2 + x^2}$	$x = a\tan\theta$	$\sec^2\theta = 1 + \tan^2\theta$

Example 7.2.8 An Exponential Integral. Determine $\displaystyle\int \frac{e^t}{(4+e^{2t})^{3/2}}\,dt$.

This is adapted from *Thomas' Calculus, 11^{th} edition, exercise 30 on page 591.*

Solution. Even though this integral involves only exponentials, we can simplify it using trigonometric substitution due to the presence of the radical. First, we'll use u-substitution to rewrite the integral into a form more suitable for trigonometric substitution: with $u = e^t$ and $du = e^t\,dt$, we get

$$\int \frac{e^t}{(4+e^{2t})^{3/2}}\,dt = \int \frac{du}{(4+u^2)^{3/2}}.$$

The appearance of $4+u^2$ in the denominator suggests that we use a tangent substitution, namely, $u = 2\tan\theta$. Then $du = 2\sec^2\theta\,d\theta$, and we get

$$\begin{aligned}\int \frac{du}{(4+u^2)^{3/2}} &= \int \frac{2\sec^2\theta}{4^{3/2}(1+\tan^2\theta)^{3/2}}\,d\theta\\ &= \frac{1}{4}\int \cos\theta\,d\theta\\ &= \frac{1}{4}\sin\theta + C\\ &= \frac{1}{4}\sin\arctan\frac{u}{2} + C\\ &= \frac{1}{4}\sin\arctan\frac{e^t}{2} + C\end{aligned}$$

Using triangles, we can simplify this further to get

$$\int \frac{e^t}{(4+e^{2t})^{3/2}}\,dt = \frac{e^t}{4\sqrt{4+e^{2t}}} + C.$$

□

7.3 Partial Fractions

At this point we have tools available to us that allow us to compute integrals of products (Section 7.1), and trigonometric functions and certain radicals(Section 7.2). Now we'll move on to a tool that will help us integrate rational functions (quotients of polynomials). The idea behind this method is to write a rational function, such as $\frac{1}{x^2-1}$, in terms of simpler fractions. We'll demonstrate by example.

Example 7.3.1 Integrating a Rational Function. Let

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

Find $\int f(x)\, dx$.

Solution. We could try secant substitution, but that we'll instead focus on simplifying this fraction. First, factor the denominator:

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

Now, assume that we can find constants A_1, A_2 such that

$$\frac{1}{x^2 - 1} = \frac{A_1}{x-1} + \frac{A_2}{x+1}.$$

If we could find such constants, we could integrate the right hand side instead of the left hand side. There are a few ways to find these constants, but the easiest might be to just clear fractions by multiplying both sides by $x^2 - 1$:

$$1 = A_1(x+1) + A_2(x-1). \tag{II.2}$$

So A_1, A_2 need to satisfy this equation *for all possible values of* x. We'll pick values for x that make it easy to find A_1 and A_2. If $x = -1$, then (II.2) becomes

$$1 = -2A_2 \Rightarrow A_2 = -\frac{1}{2}.$$

Similarly, $A_1 = \frac{1}{2}$. Therefore

$$\begin{aligned}\int \frac{1}{x^2-1}\, dx &= \int \left(\frac{1}{2}\frac{1}{x-1} - \frac{1}{2}\frac{1}{x+1}\right) dx \\ &= \frac{1}{2}\ln\left|\frac{x-1}{x+1}\right| + C.\end{aligned}$$

□

Finding partial fraction decompositions as in Example 7.3.1 isn't too bad as long as the denominator can be factored into distinct (i.e. non-repeated) *linear factors*. The situation is more complicated if the factors are repeated or are quadratics (or both).

Example 7.3.2 Finding a Partial Fraction Decomposition with Repeated Powers. Find the partial fraction decomposition for $\frac{1}{(x^2-1)^2}$.

Solution. Since $\frac{1}{(x^2-1)^2} = \frac{1}{(x-1)^2(x+1)^2}$, it's tempting to write

$$\frac{1}{(x^2-1)^2} = \frac{A_1}{x-1} + \frac{A_2}{x+1}.$$

Unfortunately, doing so gives $A_1 = A_2 = 0$, which obviously won't work. If we instead use

$$\frac{1}{(x^2-1)^2} = \frac{A_1}{(x-1)^2} + \frac{A_2}{(x+1)^2},$$

we get $A_1 = A_2 = \frac{1}{4}$. However,

$$\frac{1}{(x^2-1)^2} \neq \frac{\frac{1}{4}}{(x-1)^2} + \frac{\frac{1}{4}}{(x+1)^2}.$$

What we need to do is assume

$$\frac{1}{(x^2-1)^2} = \frac{A_1}{x-1} + \frac{A_2}{x+1} + \frac{A_3}{(x-1)^2} + \frac{A_4}{(x+1)^2}.$$

If we don't account for *each* power, we're not guaranteed a partial fraction decomposition. Now let's clear fractions:

$$1 = A_1(x-1)(x+1)^2 + A_2(x+1)(x-1)^2 + A_3(x+1)^2 + A_4(x-1)^2. \quad \text{(II.3)}$$

Quickly, using $x = -1$ we find that $A_4 = \frac{1}{4}$, and using $x = 1$ we find that $A_3 = \frac{1}{4}$.

At this point there are no nice values left to plug into the right hand side of (II.3), so we need to find A_1 and A_2 another way. One possibility is to differentiate both sides to make the right hand side simpler, which gives

$$0 = A_1(x+1)^2 + 2A_1(x^2-1) + A_2(x-1)^2 + 2A_2(x^2-1) + 2A_3(x+1) + 2A_4(x-1). \quad \text{(II.4)}$$

Now, entering $x = -1$ into (II.4) gives

$$0 = 4A_2 - 4A_4$$

and so $A_2 = A_4 = \frac{1}{4}$. Similarly, $A_1 = -A_3 = -\frac{1}{4}$. Therefore

$$\frac{1}{(x^2-1)}^2 = \frac{1}{4}\left[-\frac{1}{x-1} + \frac{1}{x+1} + \frac{1}{(x-1)^2} + \frac{1}{(x+1)^2}\right].$$

□

In order to find partial fraction decompositions when the denominator contains a quadratic, we must allow for the possibility of the numerator being *linear*.

Example 7.3.3 Partial Fraction Decomposition with a Quadratic. Determine

$$\int \frac{5x^2 + x^3}{x^5 + 3x^3}\, dx.$$

Solution. First, we can simplify:

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

Now that we've factored the denominator, we can attempt to find the partial fraction decomposition. So assume that

$$\frac{x+5}{x(x^2+3)} = \frac{A_1}{x} + \frac{A_2 x + A_3}{x^2+3}$$

Note the linear term $A_2x + A_3$ above the quadratic!

The rest of this computation goes like it did before. If we clear fractions, we get

$$x + 5 = A_1(x^2+3) + A_2x^2 + A_3x.$$

If we plug in $x = 0$, we obtain $A_1 = \frac{5}{3}$. At this point, we could plug in $x = \sqrt{3}i$ (which would be okay!) or differentiate both sides, but it might be easier to take the following approach: expand the right hand side and collect like terms to get

$$x + 5 = (A_1 + A_2)x^2 + A_3x + 3A_1.$$

For the right hand side to equal the left hand side, $A_1 + A_2$ *must be zero.* Hence $A_2 = -A_1 = -\frac{5}{3}$. Similarly, A_3 *must equal one.* Therefore

$$\begin{aligned}
\int \frac{x^3+5x^2}{x^5+3x^3}\,dx &= \int \left[\frac{5/3}{x} + \frac{1-\frac{5}{3}x}{x^2+3}\right] dx \\
&= \frac{5}{3}\int \frac{1}{x}\,dx + \int \frac{1}{x^2+3}\,dx - \frac{5}{3}\int \frac{x}{x^2+3}\,dx \\
&= \frac{5}{3}\ln|x| + \frac{1}{\sqrt{3}}\tan^{-1}\frac{x}{\sqrt{3}} - \frac{5}{6}\ln(x^2+3).
\end{aligned}$$

□

Comparing like terms as in Example 7.3.3 is a useful technique even beyond partial fractions.

The method of partial fractions works *as long as the degree of the numerator is strictly less than that of the denominator.*

Algorithm 7.3.4 Partial Fractions Algorithm. *To find the partial fraction decomposition for the rational function* $\frac{P(x)}{Q(x)}$*, do the following:*

1. *Simplify* $\frac{P(x)}{Q(x)}$ *as much as possible, using long division if necessary when* $\deg P(x) \geq \deg Q(x)$*.*
2. *Factor* $Q(x)$*.*
3. *Based on the factors of* $Q(x)$*, determine the proper form of the partial fraction decomposition.*
4. *Clear fractions.*
5. *Find the unknown constants of the decomposition by plugging values for* x *that simplify the equation, differentiating both sides, and/or comparing like terms.*

Example 7.3.5 Using Long Division. Determine $\int \frac{2y^4}{y^3-y^2+y-1}\,dy$.

Solution. Before setting up partial fractions, we should simplify the integrand as much as possible. Here, we can use long division since the degree of the numerator is greater than or equal to the degree of the denominator. Doing so, we get

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

We can integrate the first two terms on the right hand side easily, but we'll use partial fractions to rewrite the fraction:

$$\begin{aligned}
\frac{2}{y^3-y^2+y-1} &= \frac{2}{y^2(y-1)+(y-1)} \\
&= \frac{2}{y^2(y-1)} \\
&= \frac{A_1}{y} + \frac{A_2}{y^2} + \frac{A_3}{y-1}
\end{aligned}$$

Now clear fractions to get

$$2 = A_1 y(y-1) + A_2(y-1) + A_3 y^2.$$

Setting $y = 0$ gives $A_2 = -2$, and setting $y = 1$ gives $A_3 = 2$. To find A_1, we can expand the right hand side to write

$$2 = A_1 y^2 - A_1 y + A_2 y - A_2 + A_3 y^2.$$

Hence $A_1 = -A_3 = -2$.

Therefore

$$\int \frac{2y^4}{y^3 - y^2 + y - 1}\,dy = \int \left[2y + 2 - \frac{2}{y} - \frac{2}{y^2} + \frac{2}{y-1}\right] dy$$
$$= y^2 + 2y - 2\ln|y| + \frac{2}{y} + 2\ln|y-1| + C.$$

□

The next example shows a substitution that may occasionally work when partial fractions fails.

Example 7.3.6 Twisty Substitution. Determine $\int_{\frac{1}{3}}^{3} \frac{\sqrt{x}}{x^2+x}\,dx$.

Solution. Once again we have a radical causing problems, and it's not in any of the three trigonometric substitution forms we can use either. We want it to go away, so we'll try just that: set $u = \sqrt{x}$, which gives $du = \frac{1}{2\sqrt{x}}\,dx$. Then

$$\int_{\frac{1}{3}}^{3} \frac{\sqrt{x}}{x^2+x}\,dx = \int_{\frac{1}{\sqrt{3}}}^{\sqrt{3}} \frac{2x}{x^2+x}\,du.$$

Now, we use the fact that $x = u^2$ to write

$$\int_{\frac{1}{\sqrt{3}}}^{\sqrt{3}} \frac{2x}{x^2+x}\,du = \int_{\frac{1}{\sqrt{3}}}^{\sqrt{3}} \frac{2u^2}{u^4+u^2}\,du$$

which simplifies to

$$\int_{\frac{1}{\sqrt{3}}}^{\sqrt{3}} \frac{2}{u^2+1}\,du = 2\tan^{-1}(\sqrt{3}) - 2\tan^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{\pi}{3}.$$

□

7.4 Improper Integrals

(Note: this section corresponds to Section 6.6 of the text.) Consider the following problem: determine the smallest velocity required for an object to escape the earth's gravitational pull (i.e., the **escape velocity**). We can try to answer this problem by looking at the total work W needed to move this object. From one viewpoint, work is just the total change in kinetic energy. If the object is moving at escape velocity, say v_e, then all of its kinetic energy must have been converted into work: $W = \frac{1}{2}m_O v_e^2$, where m_O is the mass of the object.

As we'll see later, work is also the integral of force, so let's examine the force on the object. If we neglect air resistance and assume that the only force acting on the object is gravity, then the force exerted on the object is given by

$$F(r) = \frac{m_E m_O}{r^2} G,$$

where m_E is the mass of the Earth, m_O is still the mass of the object, r is the distance from the center of mass of the Earth to the center of mass of the object, and G is the gravitational constant. Now assume that the Earth is a sphere of radius R_0, and that the object has enough velocity to move it a distance of R_1 units from the surface of the Earth. Then the total work done in moving the object this distance is

$$\int_{R_0}^{R_1} F(r)\,dr = m_E m_O G \int_{R_0}^{R_1} \frac{1}{r^2}\,dr.$$

Now let's think about should happen if the object is at escape velocity. In this case, theoretically at least, there should be no limit to the distance the object can travel if given enough time. So what's the total work done?

Let's try to compute $m_E m_O G \int_{R_0}^{R_1} \frac{1}{r^2}\,dr$. We get

$$\begin{aligned} m_E m_O G \int_{R_0}^{R_1} \frac{1}{r^2}\,dr &= -\frac{m_E m_O G}{r}\Big]_{R_0}^{R_1} \\ &= \frac{m_E m_O G}{R_0} - \frac{m_E m_O G}{R_1}. \end{aligned}$$

Note that this is still an underestimate for the total work theoretically done, since R_1 can increase without bound. So let's send R_1 to ∞ and see what we get:

$$\lim_{R_1\to\infty} \frac{m_E m_O G}{R_0} - \frac{m_E m_O G}{R_1} = \frac{m_E m_O G}{R_0}.$$

So, to summarize, the total work W should be

$$W = \lim_{R_1\to\infty}\left(m_E m_O G \int_{R_0}^{R_1} \frac{1}{r^2}\,dr\right) = \frac{1}{2} m_O v_e^2$$

or just

$$W = \frac{m_E m_O G}{R_0} = \frac{1}{2} m_O v_e^2.$$

We can solve this equation for v_e, and so the escape velocity should be

$$v_e = \sqrt{\frac{2 m_E G}{R_0}}.$$

As a neat bonus, this shows that the escape velocity only depends on G, the mass of the Earth m_E and the "radius" of the Earth R_0. Plugging in values for these figures, we get $v_e \approx 11185$ meters per second, or 11.185 kilometers per second.

This is all very important obviously, but this is calculus, and what *we* really care about are the mathematical tools required to solve this problem. To do this, we basically had to figure out what the integral of $\frac{1}{r^2}$ was over the interval $[R_0, \infty)$. This is our first example of an **improper integral**.

Improper Integrals over Infinite Intervals

First, a definition.

Definition 7.4.1 Type 1 Improper Integral. Let $f(x)$ be some function, and let a be a constant. If $\int_a^b f(x)\,dx$ exists for every $b \geq a$, then we define

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx,$$

assuming this limit exists. Similarly, if $\int_b^a f(x)\,dx$ exists for every $b \leq a$, then we define

$$\int_{-\infty}^{a} f(x)\,dx = \lim_{b\to-\infty} \int_{b}^{a} f(x)\,dx,$$

assuming this limit exists. These improper integrals are **convergent** if the corresponding limits exist and **divergent** otherwise.

Finally, if both $\int_{-\infty}^{a} f(x)\,dx$ and $\int_{a}^{\infty} f(x)\,dx$ are convergent, then we define

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx + \int_{a}^{\infty} f(x)\,dx.$$

◊

Example 7.4.2 A Divergent Integral. Determine if $\int_{1}^{\infty} \frac{1}{x}\,dx$ is convergent or divergent.

Solution. By definition, we have

$$\begin{aligned}\int_{1}^{\infty} \frac{1}{x}\,dx &= \lim_{b\to\infty} \int_{1}^{b} \frac{1}{x}\,dx \\ &= \lim_{b\to\infty} (\ln b - \ln 1) \\ &= \infty.\end{aligned}$$

Hence the integral is divergent. □

In Example 7.4.2, note that $\frac{1}{x} \to 0$ as $x \to \infty$. However, it doesn't go to 0 *fast enough* for the integral to converge.

Example 7.4.3 Is $\int_{1}^{\infty} \frac{1}{x^2}\,dx$ convergent? □

The convergence of $\int_{1}^{\infty} \frac{1}{x^p}\,dx$ depends on the value of p.

Theorem 7.4.4 Integral p-test. *The integral $\int_{1}^{\infty} \frac{1}{x^p}\,dx$ converges if and only if $p > 1$.*

Example 7.4.5 Area Under a Graph. Find the area under $f(x) = xe^{-x^2}$.

Solution. The area under $f(x)$ is given by $\int_{-\infty}^{\infty} xe^{-x^2}$. To evaluate this integral, we must split it into two improper integrals:

$$\begin{aligned}\int_{-\infty}^{\infty} xe^{-x^2}\,dx &= \int_{-\infty}^{0} xe^{-x^2}\,dx + \int_{0}^{\infty} xe^{-x^2}\,dx \\ &= \lim_{a\to\infty} \int_{a}^{0} xe^{-x^2}\,dx + \lim_{b\to\infty} \int_{0}^{b} xe^{-x^2}\,dx\end{aligned}$$

At this point, we can use u-substitution to find each integral, using $u = -x^2$ and $du = -2x\,dx$. So we get

$$\begin{aligned}\lim_{a\to\infty} \int_{a}^{0} xe^{-x^2}\,dx + \lim_{b\to\infty} \int_{0}^{b} xe^{-x^2} &= \lim_{a\to\infty} \int_{-a^2}^{0} \frac{-e^u}{2}\,du + \lim_{b\to\infty} \int_{0}^{-b^2} \frac{-e^u}{2}\,du \\ &= \lim_{a\to-\infty} \left[-\frac{1}{2} + \frac{e^{-a^2}}{2}\right] + \lim_{b\to\infty} \left[-\frac{e^{-b^2}}{2} + \frac{1}{2}\right] \\ &= \lim_{a\to\infty} \frac{e^{-a^2}}{2} - \lim_{b\to\infty} \frac{e^{-b^2}}{2} \\ &= 0.\end{aligned}$$

Hence the area under $f(x)$ is 0 (which might not be so surprising if you graph $f(x)$). □

Example 7.4.6 Radioactive Decay. A radioactive substance undergoes decay, and has mass $m(t)$ at time t given by

$$m(t) = m_0 e^{-kt}$$

for some $k > 0$. The *expected lifetime* of a particle is given by

$$k\int_0^\infty te^{-kt}\,dt.$$

Find this value.

Solution. We can solve this using integration by parts. Doing so, we get

$$\begin{aligned} k\int_0^\infty te^{-kt}\,dt &= -\lim_{b\to\infty}\left[te^{-kt} + \frac{1}{k}e^{-kt}\right]_0^b \\ &= \frac{1}{k} \end{aligned}$$

This can be verified using the code cell below. □

```
from sympy import *
init_printing()

# We need to assume that k is positive for the integral to
    converge.
# This will also make SymPy assume that t is positive, but
    that's not a problem for us
# since we're integrating over [0,oo).
k,t = symbols('k t', positive=True)
m = k*t*exp(-k*t)

m.integrate((t, 0, oo))
```

Example 7.4.7 A Probability Distribution. In probability, a **probability distribution** is a function $f(x)$ satisfying

$$\int_{-\infty}^\infty f(x)\,dx = 1.$$

Is $f(x) = \int_{-\infty}^\infty \frac{1}{1+x^2}\,dx$ a probability distribution?

Solution. We have

$$\begin{aligned} \int_{-\infty}^\infty \frac{1}{1+x^2}\,dx &= \tan^{-1}x\Big]_{-\infty}^\infty \\ &= \lim_{b\to\infty}\tan^{-1}x\Big]_0^b + \lim_{c\to-\infty}\tan^{-1}x\Big]_c^0 \\ &= \pi. \end{aligned}$$

So this isn't a probability distribution. □

Example 7.4.8 Another Laplace Transform. The **Laplace transform** of a function $f(t)$ is defined to be

$$\int_0^\infty f(t)e^{-st}\,st,$$

assuming s is chosen so that the integral converges. Find the Laplace transform of $f(t) = t$. □

Improper Integrals with Discontinuous Integrands

The second type of improper integral we'll consider involves integrands with "divide-by-zero" problems.

Definition 7.4.9 Type 2 Improper Integral. If $f(x)$ is continuous on $[a, b)$ and is discontinuous at b, then

$$\int_a^b f(x)\,dx = \lim_{c\to b^-} \int_a^c f(x)\,dx$$

assuming this limit exists.

If $f(x)$ is continuous on $(a, b]$ and is discontinuous at a, then

$$\int_a^b f(x)\,dx = \lim_{c\to a^+} \int_c^b f(x)\,dx$$

assuming this limit exists.

If $f(x)$ is at c where $a < c < b$, then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

assuming these integrals are convergent. ◊

Example 7.4.10 Logarithmic Discontinuity. Determine if $\int_0^1 \ln x\,dx$ is convergent.

Solution. This integral is improper since $\ln x$ is discontinuous at $x = 0$. To evaluate it, we'll use limits:

$$\begin{aligned}\int_0^1 \ln x\,dx &= \lim_{c\to 0^+} \int_c^1 \ln x\,dx \\ &= \lim_{c\to 0^+} [x\ln x - x]_c^1 \\ &= -1 - \lim_{c\to 0^+} c\ln c \\ &= -1,\end{aligned}$$

where we used the fact that $c\ln c \to 0$ as $c \to 0$ by L'Hopital's Rule. □

Example 7.4.11 Another Integral p-Test. For what values of p is the integral $\int_0^1 \frac{1}{\sqrt[p]{x}}\,dx$ convergent?

Solution. We could try to compute this as is, but we can save ourselves some work by making use of Theorem 7.4.4. First, we'll use the substitution $u = x^{-\frac{1}{p}}$. Then $du = -\frac{1}{p}x^{-\frac{1}{p}-1}$ and

$$\begin{aligned} u \to \infty \quad &\text{as} \quad x \to 0^+ \\ u = 1 \quad &\text{if} \quad x = 1. \end{aligned}$$

Therefore

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt[p]{x}}\,dx &= \int_\infty^1 u\left[-pu^{-1-p}\right]\,du \\ &= p\int_1^\infty \frac{1}{u^p}\,du\end{aligned}$$

which we know converges if and only if $p > 1$. Hence the original integral converges if and only if $p > 1$. □

Comparison Tests for Improper Integrals

For integrals, we have two comparison tests that allow us to determine if an improper integral converges by comparing it with a simpler integral. First, we'll look at the **direct comparison test**.

Theorem 7.4.12 Direct Comparison Test. *Suppose that $f(x)$ and $g(x)$ are continuous functions and that*

$$f(x) \geq g(x) \geq 0$$

for $x \geq a$, for some real number a. Then if $\int_a^\infty f(x)\,dx$ converges, so goes $\int_a^\infty g(x)\,dx$. Likewise, if $\int_a^\infty g(x)\,dx$ diverges, then so does $\int_a^\infty f(x)\,dx$.

Example 7.4.13 Does $\int_1^\infty \frac{1}{\sqrt{e^x+x}}\,dx$ converge?

Solution. This integrand looks pretty awful. However, we can say that

$$\frac{1}{\sqrt{e^x+x}} \leq \frac{1}{\sqrt{e^x}} = e^{-\frac{x}{2}}.$$

Since

$$\int_1^\infty e^{-\frac{x}{2}}\,dx = 2\sqrt{e},$$

Theorem 7.4.12 implies that the original integral must converge as well. Note that we don't know *what* it converges to, only that it does. □

Example 7.4.14 The Gaussian Integral. Show that the integral $\int_0^\infty e^{-x^2}\,dx$ is convergent.

Solution. We can do this by breaking the integral into two parts: $\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx$. If x is between 0 and 1, then $e^{-x^2} \leq e^0 = 1$. On the other hand, if $1 \leq x < \infty$, then $e^{-x^2} \leq e^{-x}$. Therefore

$$\begin{aligned}\int_0^\infty e^{-x^2}\,dx = \int_0^\infty e^{-x^2}\,dx &= \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx \\ &\leq \int_0^1 dx + \int_1^\infty e^{-x}\,dx \\ &= 1 + \frac{1}{e} \\ &< \infty.\end{aligned}$$

Hence the original integral must also be convergent. It's value is actually known, and SymPy can be used to compute the integral, as below. □

```
# Computing the Gaussian with SymPy.
import sympy as sp
sp.init_printing()

x = sp.Symbol('x')
sp.integrate(sp.exp(-x**2), (x, 0, oo))
```

To use Theorem 7.4.12 successfully, we need to choose a simpler function that's related to the integrand and then use algebra to justify a specific inequality. It is sometimes more straightforward to apply the **limit comparison test**.

Theorem 7.4.15 Limit Comparison Test. *Let $f(x)$ and $g(x)$ be positive, continuous functions on $[a, \infty)$ for some constant a. If*

$$\lim_{x\to\infty} \frac{f(x)}{g(x)}$$

exists and is positive, then the integrals

$$\int_a^\infty f(x)\,dx \quad \text{and} \quad \int_a^\infty g(x)\,dx$$

must both converge or both diverge.

Example 7.4.16 Determine if $\displaystyle\int_1^\infty \frac{3x^3+2x^5}{5x^7+4x^2}\,dx$ converges.

Solution. The integrand is awful again, but if x is very large then we can say that

$$\frac{3x^3+2x^5}{5x^7+4x^2} \approx \frac{1}{x^2}.$$

Since $\int_1^\infty \frac{1}{x^2}\,dx$ converges, this suggests that maybe the original integral does as well. We can prove this using Theorem 7.4.15.

In fact, we have

$$\lim_{x\to\infty} \frac{\frac{3x^3+2x^5}{5x^7+4x^2}}{\frac{1}{x^2}} = \frac{2}{5}.$$

Since this limit exists and is positive, then the original integral must also converge by the Limit Comparison Test. □

Chapter 8

Applications of Integration

8.1 Areas Between Curves

We already know that $\int_a^b f(x)\,dx$ geometrically represents the area under $f(x)$ from $x = a$ to $x = b$. Now we'd like to try to extend the integral to other geometrical concepts such as surface areas and volumes. An easy generalization we can make right now is that of *areas between curves*, as in Figure 8.1.1.

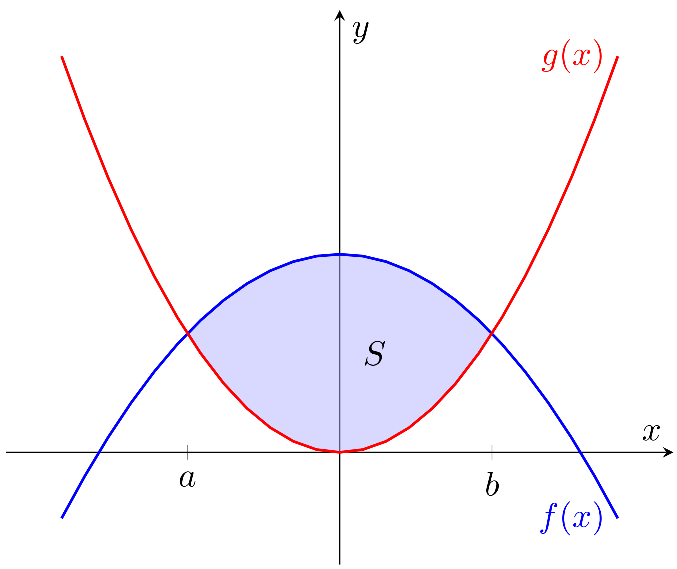


Figure 8.1.1 Area between curves. Adapted from here[2].

It may not be too much of a stretch to say that the area S should be given by $\int_a^b [f(x) - g(x)]\,dx$, and indeed it is, but we will prove this using Riemann sums for the reason that Riemann sums appear *a lot* in applications. Furthermore, Riemann sums provide us with the formula we need to evaluate integrals numerically using libraries such as NumPy. So let's start by partitioning the interval $[a, b]$ into n *subintervals*

$$[x_{i-1}, x_i] \quad \text{for} \quad 0 \le i \le n,$$

where $x_0 = a, x_n = b$ and

$$x_i - x_{i-1} = \Delta x = \frac{b-a}{n}.$$

[2] `tex.stackexchange.com/questions/164773/graphics-area-between-curves`

Now we'll use the subintervals to break the region S up into rectangles to approximate the area. So choose a point x_i^* from each subinterval $[x_{i-1}, x_i]$ and use it to determine the height of the corresponding rectangle: $f(x_i^*) - g(x_i^*)$. Then the area is $f(x_i^*) - g(x_i^*)\Delta x$, and adding up the areas of all of the rectangles lets us approximate the area of S:

$$\text{area of } S \approx \sum_{i=1}^{n} [f(x_i^*) - g(x_i^*)]\Delta x \tag{II.1}$$

Now send $n \to \infty$ (equivalently, send $\Delta x \to 0$) to make the approximation exact: area of $S = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i^*) - g(x_i^*)\Delta x = \int_a^b [f(x) - g(x)]\, dx$.

Example 8.1.2 Area Between Parabolas. Find the area between $f(x) = x^2 + 1$ and $g(x) = 2x^2$.

Solution. If we graph f and g, we see that the area will be given by $\int_a^b [(x^2 + 1) - 2x^2]\, dx$, where a and b are the values of x where $f(x) = g(x)$. Using a bit of algebra, we get $a = -1$ and $b = 1$. Therefore the area between the parabolas is

$$\begin{aligned}\int_{-1}^{1} (1 - x^2)\, dx &= 2\int_0^1 (1 - x^2)\, dx \\ &= \frac{4}{3}.\end{aligned}$$

□

Although we found an exact solution in Example 8.1.2, we can also use this to test our NumPy skills. (II.1) shows us how we can use NumPy to find this quantity.

```
import numpy as np

# Endpoints of interval and number of subintervals
a = -1
b = 1
n = 10

# Quick way to define f(x) and g(x)
f = lambda x: x**2 + 1
g = lambda x: 2 * x**2

# Partition interval
X = np.linspace(a, b, n)

# Area of each rectangle
deltaX = (b - a) / n
area = (f(X) - g(X)) * deltaX

# Add areas together to get approximation
print(np.sum(area))
```

```
1.1851851851851853
```

Example 8.1.3 Area with Changing Limits. Find the area of the region bounded by $x = y^4$ and $y^2 + x = 2$.

Solution. If we solve both equations for y, we get $y = \pm\sqrt[4]{x}$ and $y = \pm\sqrt{2 - x}$. We can then graph the region below:

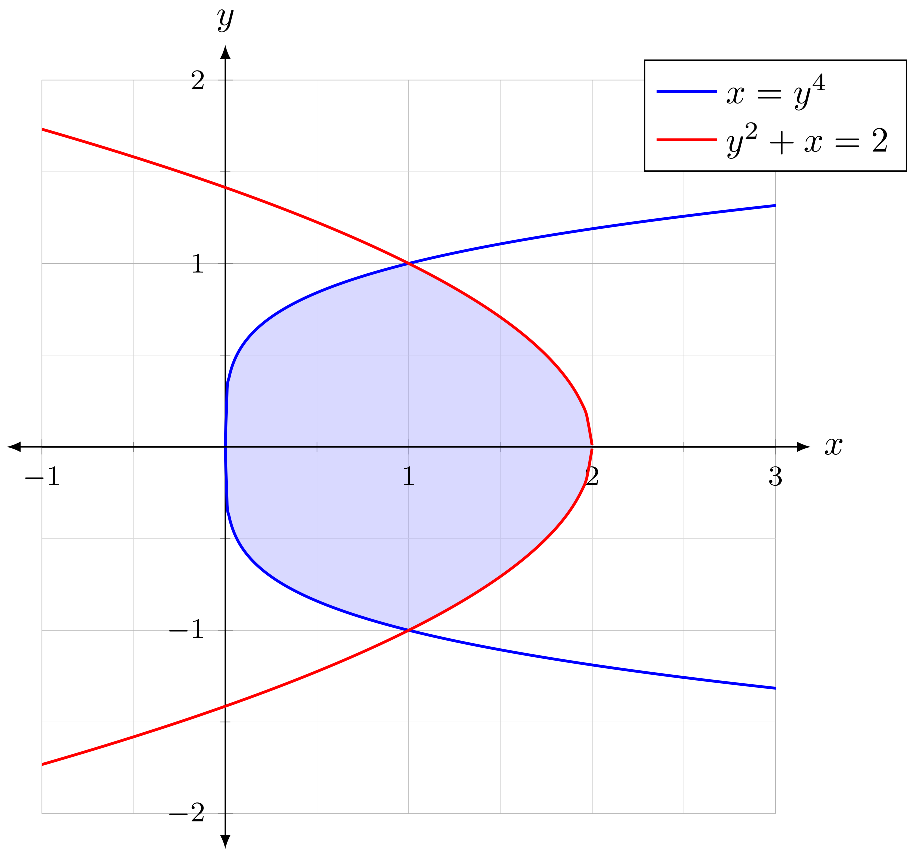


Figure 8.1.4 The region bounded by $x = y^4$ and $y^2 + x = 2$.

To find the area between these curves, we'll need to compute *two* integrals: one for $0 \leq x \leq 1$ and another for $1 \leq x \leq 2$. In particular, the area will be given by $\int_0^1 [x^{\frac{1}{4}} - (-x^{\frac{1}{4}})]\, dx + \int_1^2 [\sqrt{2-x} - (-\sqrt{2-x})]\, dx$. Doing so, we get $\frac{44}{15}$ for the area between the curves. □

Part of the reason why finding the area in Example 8.1.3 was more complicated was because the curves did not represent functions of x. However, we can view each curve as representing a function of y, namely, $x(y) = y^4$ and $x(y) = 2 - y^2$. Then the area can be found as $\int_{-1}^{1} [(2-y^2) - y^4]\, dy = \frac{44}{15}$. So it can be much easier if we find the area by integrating with respect to y as opposed to x.

Example 8.1.5 Find the area contained by the curves $x = y^2 - 4y$ and $x = 2y - y^2$.

Solution. If we graph the region, we get Figure 8.1.6.

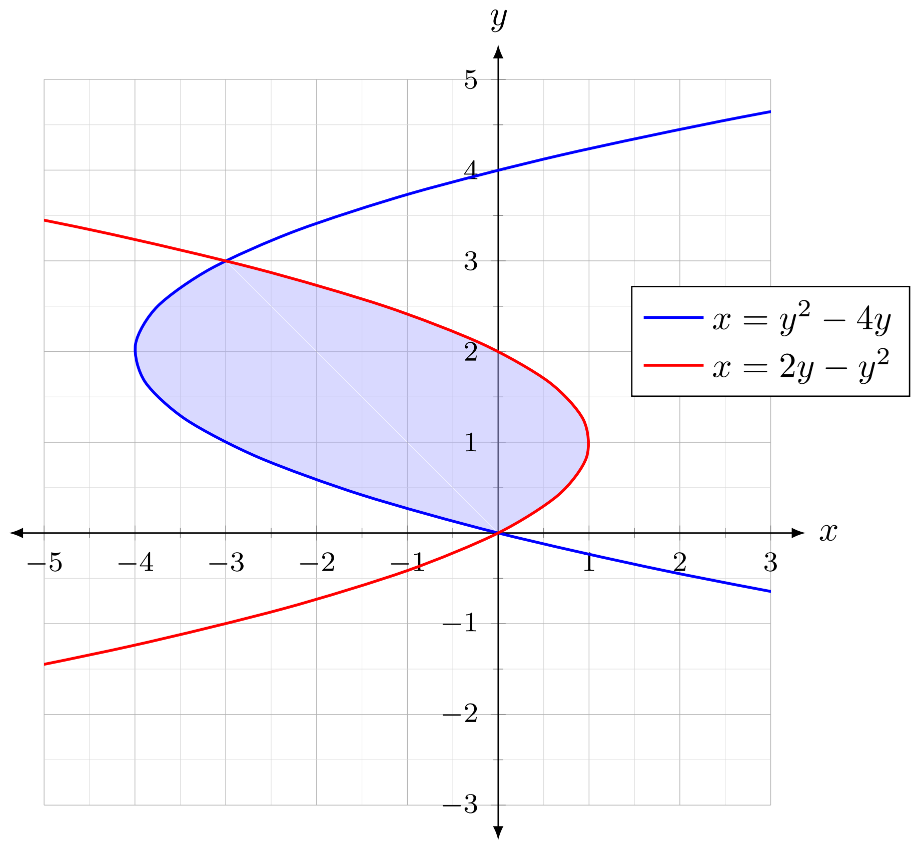


Figure 8.1.6 Region between two parabolas.

Finding the area by using integrals with respect to x would require setting up and solving three different integrals! So, not terrific. However, we can integrate with respect to y fairly easily. By finding where the curves intersect, we see that $0 \le y \le 3$. Hence the area should be

$$\begin{aligned}\int_0^3 \left[(2y - y^2) - (y^2 - 4y)\right] dy &= \int_0^3 \left[6y - 2y^2\right] dy \\ &= 3y^2 - \frac{2}{3}y^3 \Big]_0^3 \\ &= 27 - 18\end{aligned}$$

and so the area is just 9 units. □

8.2 Volumes

Just as we can with areas, we can attempt to find volumes using integrals as well. At the moment we can't find volumes of general regions without first describing integrals in higher dimensions, which is the content of Chapter 13. Until then, we'll need to restrict ourselves to so-called **solids of revolution**.

As an example, consider Figure 8.2.1. To generate the solid of revolution on the right, we take the graph of $y = f(x)$ and rotate it about the x-axis, creating a three-dimensional region.

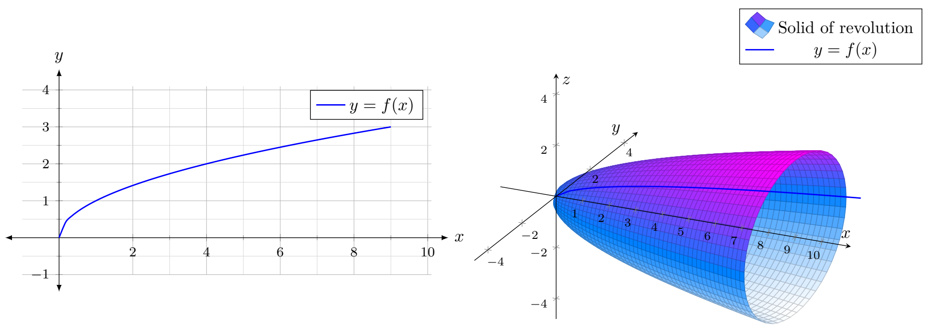


Figure 8.2.1 Generating a solid of revolution.

To find the volume of such a region, the idea is to find the *area* of a general cross-section of the region, which by design should be a function of x alone, say $A(x)$. Then, roughly, $A(x)\Delta x$ should represent the volume of a small cross-section of the solid, and so the integral $\int_a^b A(x)\,dx$ should give us the volume of the region. This entire argument can be made precise using Riemann sums as in the previous section, and so we get the following: the solid S from $x = a$ to $x = b$ whose cross-sections (with respect to the x-axis) have area $A(x)$ has volume given by

$$V = \int_a^b A(x)\,dx. \tag{II.2}$$

Example 8.2.2 Finding the Volume of a Sphere. Find the volume of a sphere of radius r.

Solution. To begin, let's assume that the sphere is centered at the origin. Then each cross section perpendicular to the x-axis is just a circle with radius $\sqrt{r^2 - x^2}$. Hence each cross-sectional area is given by $A(x) = \pi(r^2 - x^2)$, which means the volume of the sphere should be

$$\begin{aligned}\int_{-r}^{r} \pi(r^2 - x^2)\,dx &= 2\int_0^r \pi(r^2 - x^2)\,dx \\ &= 2\left[\pi r^2 x - \frac{\pi x^3}{3}\right]_0^r \\ &= 2\left[\pi r^3 - \frac{\pi r^3}{3}\right] \\ &= \frac{4}{3}\pi r^3.\end{aligned}$$

□

Again, the idea for finding the volumes of these regions is to determine the cross-sectional areas perpendicular to the x-axis (or whatever the axis of rotation happens to be) and then integrate.

Example 8.2.3 Find the volume of the region obtained by rotating the graph of $y = \sqrt{x}$ from $x = 1$ to $x = 4$ about the line $y = 2$.

Solution. As is almost always the case, a good way to start is by graphing the region to get a rough idea of what it looks like. Doing so, we see that each of the cross-sections are circles. But this is to be expected, since the region is a solid of revolution. Given x between 1 and 4, the area of the corresponding cross-section is given by

$$A(x) = \pi(2 - \sqrt{x})^2.$$

Hence the volume of this region should be

$$\begin{aligned}\int_1^4 \pi(2-\sqrt{x})^2\,dx &= \pi\int_1^4 (2-4\sqrt{x}+x)\,dx \\ &= \pi\left[2x-\frac{8}{3}x^{\frac{3}{2}}+\frac{x^2}{2}\right]_1^4 \\ &= \pi\left[\left(8-\frac{64}{3}+8\right)-\left(2-\frac{8}{3}+\frac{1}{2}\right)\right]\end{aligned}$$

□

The previous problems involved finding volumes by using disks since disks were natural cross-sections. The next example instead uses washers as cross-sections instead. Although the formula we get for the volume is slightly different, it's still essentially integrating the areas of the cross-sections.

Example 8.2.4 Find the volume of the solid obtained by rotating the region bounded by $y = x$, $y = \sqrt{x}$ and $y = 1$ about the line $x = 2$.

Solution. Since the axis of rotation is $x = 2$, we want to look at cross-sections perpendicular to this line, i.e., parallel to the x-axis. Each such cross-section will look like a washer, and hence its area will involve the inner radius and outer radius of the washer. Since the radii are measured in terms of horizontal distance, we'll need to find x in terms of y. Doing so, we see that the inner radius is given by $2 - y$ while the outer radius is given by $2 - y^2$. Therefore the volume of the region is given by

$$\begin{aligned}\int_0^1 \left[\pi(2-y^2)^2-\pi(2-y)^2\right]\,dy &= \pi\int_0^1 \left[(4-4y^2+y^4)-(4-2y+y^2)\right]\,dy \\ &= \pi\int_0^1 \left[y^4-5y^2+2y\right]\,dy\end{aligned}$$

□

8.3 Volumes by Cylindrical Shells

The method we used to find volumes in Section 8.2 applies to solids that have cross-sections depending on x or y alone. Once we had a formula for this area, we could then integrate it to get the volume. However, not all shapes have such simple cross-sections, and so this method doesn't apply in general. In this section, we'll look at using cylindrical shells instead.

Therefore, the following fact will prove useful: the volume of a cylindrical shell of height h, inner radius r_1 and outer radius r_2 is given by

$$V = 2\pi\frac{r_1+r_2}{2}h(r_2-r_1) = 2\pi rh\Delta r \tag{II.3}$$

where r is the "average radius" and Δr is the thickness of the shell. If we can break a solid up into these cylindrical "cross-sections," then integrating these areas should again give us the volume of the solid. For shells that are perpendicular to the x-axis, we can typically replace r with x and h with $f(x)$, leading to an integral of the form

$$\int_a^b 2\pi x f(x)\,dx.$$

Finding areas using disks and washers usually involves finding disks/washers that are perpendicular to the axis of rotation. When finding volumes using shells, we'll typically use shells that are *parallel* to the axis of rotation.

One final note: this method, as in Section 8.2, applies only to solids with a certain degree of symmetry. If you continue on to Calculus 3, the techniques you'll learn in Chapter 13 will replace this method as they do the disk and washer methods from Section 8.2. In particular, this method is a special case of the transformation to *cylindrical coordinates* presented in Section **??**.

Example 8.3.1 Finding Volumes Using Cylindrical Shells. Find the volume of the solid obtained by rotating the region bounded by $y = x^2$, $y = 0$ and $x = 1$ about the y-axis.

Solution. The axis of rotation is the y-axis, so each shell should be perpendicular to this axis as well. Drawing a typical shell, we see that the radius at x should be x while the height is $f(x) = x^2$. So the volume of this region should be

$$\int_0^1 2\pi x(x^2)\,dx = \frac{\pi}{2}.$$

□

It should be noted that the volume in Example 8.3.1 could also have been found using washers instead of shells. If you can, you should use disks or washers before using shells, since integrals derived from using shells are typically more complicated.

Example 8.3.2 Find the volume of the solid obtained by rotating the region bounded by $x = 0$, $y = 4x - x^2$ and $y = 8x - 2x^2$ about the line $x = -2$.

Solution. A good first step, as always, is to plot the region. Since $y = 4x - x^2$ and $y = 8x - 2x^2$ give downward opening parabolas, we can quickly sketch a rough graph by finding the x-coordinates of the vertices, which occur at $x = 2$. Finding the corresponding y values, and noticing that each parabola intersects the x-axis at $x = 0, 4$, gives a pretty good picture of the graph.

If we tried to find the volume using disks/washers, we'd need to solve $y = 4x - x^2$ and $y = 8x - 2x^2$ for x since each cross-section would be parallel to the x-axis. This is doable using the quadratic formula, but the algebra would get very bad very quickly once we start integrating. So we'll try shells instead.

A typical shell for this region has radius $x - (-2) = x + 2$ and height $(8x - 2x^2) - (4x - x^2) = 4x - x^2$. Hence the volume is given by

$$\begin{aligned}
\int_0^4 2\pi(x+2)(4x-x^2)\,dx &= 4\pi\int_0^2 (x+2)(2x-x^2)\,dx \\
&= 4\pi\int_0^2 \left[4x - x^3\right]\,dx \\
&= 4\pi\left[2x^2 - \frac{x^4}{4}\right]_0^2 \\
&= 16\pi
\end{aligned}$$

□

8.4 Arc Length

Say the motion of a particle is described by $y = f(x)$ for $a \le x \le b$. Then the path that the particle travels will look like a curve, say

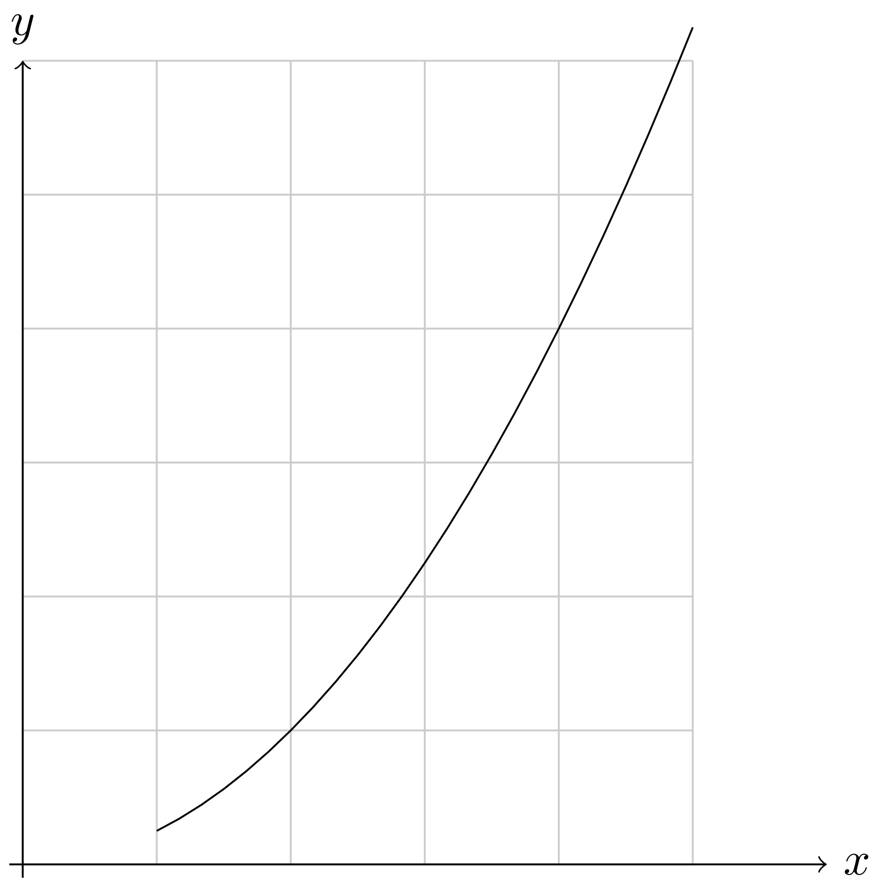


Figure 8.4.1 Trajectory of our particle.

Our goal now is to determine how far the particle travels as x goes from a to b. In other words, we want to find the length of this curve from $x = a$ to $x = b$.

If the curve were just a line we could easily find its length by using the Pythagorean theorem, but for a more general curve like this we're stuck. So we'll start by approximating the curve by using line segments, which we'll call L_1, L_2, and so on up to L_n, then we'll find the lengths of the line segments. Say the first line segment goes from $x = x_0$ to $x = x_1$, the second from $x = x_1$ to $x = x_2$ and so on. To make life simpler, we'll assume that all these intervals have equal length Δx. That is,

$$x_i - x_{i-1} = \Delta x$$

for $0 \leq i \leq n$. Also, we'll define $\Delta y_i = f(x_i) - f(x_{i-1})$.

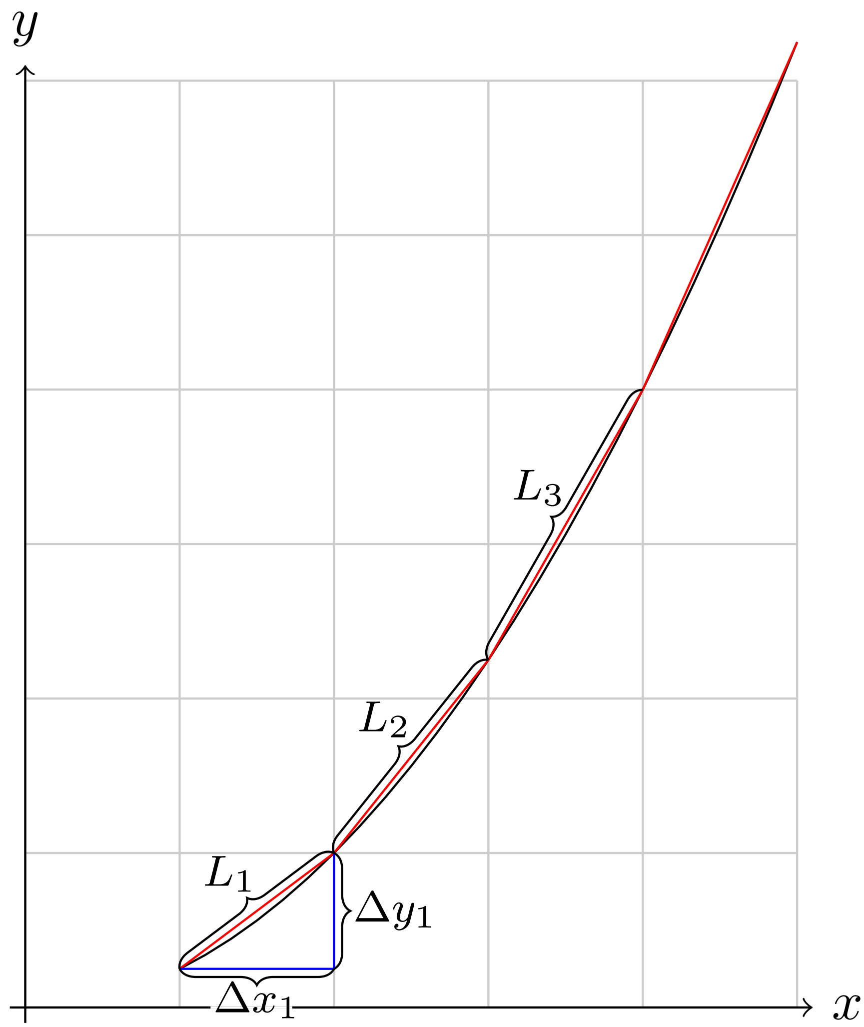


Figure 8.4.2 Approximating arc length with line segments.

If we look at L_1, then the line goes Δx in the x-direction and Δy_1 in the y-direction, so by the Pythagorean theorem,

$$\text{length of } L_1 = \sqrt{(\Delta x)^2 + (\Delta y_1)^2} = \sqrt{1 + \left(\frac{\Delta y_i}{\Delta x}\right)}\Delta x.$$

Now if $\Delta x \approx 0$, then $\frac{\Delta y_1}{\Delta x} \approx f'(x_1)$. Hence

$$\text{length of } L_1 \approx \sqrt{1 + f'(x_1)^2}\Delta x$$

if $\Delta x \approx 0$.

We can extend this estimate to the other line segments as well, so we can say that the total length L of the curve is approximately

$$\sum_{i=1}^{n} \sqrt{1 + f'(x_i)^2}\Delta x.$$

If the curve is nice enough, this approximation should become exact as $\Delta x \to 0$. So we can take this as the definition of the arc length of the curve $y = f(x)$.

Definition 8.4.3 Arc Length of a Curve. Let $y = f(x)$ denote a continuous function defined on some interval $[a, b]$. The **arc length** of $f(x)$ over this interval is defined to be

$$L = \lim_{\Delta x \to 0} \sum_{i=1}^{n} \sqrt{1 + f'(x_i)^2} \Delta x = \int_a^b \sqrt{1 + f'(x)^2} \, dx.$$

◊

Example 8.4.4 The Circumference of a Circle. Find the circumference of a circle of radius r.

Solution. To find the circumference of this circle, we need to find an equation whose graph is the circle, or at least part of it. One choice is $y = \sqrt{r^2 - x^2}$ for $-r \leq x \leq r$. Since this gives the top half of a circle, the circumference is then

$$\begin{aligned} 2 \int_{-r}^{r} \sqrt{1 + [y']^2} \, dx &= 2 \int_{-r}^{r} \sqrt{1 + \frac{x^2}{r^2 - x^2}} \, dx \\ &= 2r \int_{-r}^{r} \frac{1}{\sqrt{r^2 - x^2}} \, dx \\ &= 2r \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{r \cos \theta}{r \cos \theta} \, d\theta \\ &= 2\pi r. \end{aligned}$$

□

We could try to use Definition 8.4.3 to find the arc length of an ellipse, but it turns out the resulting integral is too complicated to have a "closed form" solution.

We can use Definition 8.4.3 to define the **arc length function**:

$$s(x) = \int_a^x \sqrt{1 + [f'(t)]^2} \, dt.$$

This function can be useful to *parameterize* movement along a curve, which you'll see in Calculus 3.

Example 8.4.5 Find the arc length function for $y = x^2$ assuming that $x = 0$ is the starting point.

Solution. We have

$$s(x) = \int_1^x \sqrt{1 + 4t^2} \, dt.$$

To find this, we'll use trigonometric substitution with $2t = \tan \theta$ to get

$$\begin{aligned} s(x) &= \frac{1}{2} \int_0^{\tan^{-1}(2x)} \sqrt{1 + \tan^2 \theta} \sec^2 \theta \, d\theta \\ &= \frac{1}{4} \left[\sec \theta \tan \theta + \ln(\sec \theta + \tan \theta) \right]_0^{\tan^{-1}(2x)} \\ &= \frac{x}{2} \sqrt{1 + 4x^2} + \frac{1}{4} \ln(\sqrt{1 + 4x^2} + 2x). \end{aligned}$$

□

8.5 Area of a Surface of Revolution

Consider a function $y = f(x)$. We generate a solid of revolution by rotating the graph of $f(x)$ from $x = a$ to $x = b$ about the x-axis. We want to find the surface area of this figure. We can't find it directly yet, but we can approximate it by using conical frustums.

The surface area of a conical frustum is given by the formula

$$S_F = \pi(R_1 + R_2)\sqrt{(\Delta x)^2 + (\Delta y)^2} = \pi(R_1 + R_2)\sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}\,\Delta x.$$

If $\Delta x \approx 0$ then we can say

$$S_F \approx 2\pi f(x)\sqrt{1 + f'(x)^2}\Delta x$$

for some $f(x)$ between R_1 and R_2.

By using multiple frustums, we can approximate the surface area of the solid. Break $[a, b]$ into n subintervals

$$[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]$$

with $x_0 = a$ and $x_n = b$, where each subinterval has length Δx. Then if S is the surface area of the solid, we have

$$\begin{aligned} S &\approx \text{sum of areas of frustums} \\ &\approx \sum_{k=1}^{n} 2\pi f(x_i)\sqrt{1 + f'(x_i)^2}\Delta x. \end{aligned}$$

This approximation should become more exact as $\Delta x \to 0$, so we have

$$S = \lim_{\Delta x \to 0} \sum_{k=1}^{n} 2\pi f(x_i)\sqrt{1 + f'(x_i)^2}\Delta x$$

or

$$S = \int_a^b 2\pi f(x)\sqrt{1 + f'(x)^2}\,dx.$$

Similarly, if we generate a solid of revolution by rotating $x = g(y)$ from $y = c$ to $y = d$ about the y-axis, the surface area of this solid is given by

$$S = \int_c^d 2\pi g(y)\sqrt{1 + g'(y)^2}\,dy.$$

Example 8.5.1 Finding Surface Area of a Sphere. Find the surface area of a sphere of radius r.

Solution. We can create a sphere of radius r by rotating the semicircle $y = \sqrt{r^2 - x^2}$ about the x-axis, from $x = -r$ to $x = r$. So $f(x) = \sqrt{r^2 - x^2}$ and $f'(x) = \dfrac{-x}{\sqrt{r^2 - x^2}}$. So the surface area is

$$S = \int_{-r}^{r} 2\pi\sqrt{r^2 - x^2}\sqrt{1 + \frac{x^2}{r^2 - x^2}}\,dx$$

or just

$$S = \int_{-r}^{r} 2\pi r\,dx.$$

□

Example 8.5.2 Find the surface area of the solid generated by rotating the graph of $y = x^2$ from $y = 0$ to $y = 3$ about the y-axis.

Solution. Since we are rotating about the y-axis, we need to use the other formula. Since $y = x^2$, we can rewrite this as $x = sqrty$. So the surface area is

$$\int_0^3 2\pi\sqrt{y}\sqrt{1+\frac{1}{4y}}\,dy.$$

□

Example 8.5.3 Find the integral that gives the surface area of the solid generated by rotating the graph of $y = x^2$ from $x = 0$ to $x = 3$ about the x-axis.

Solution. By the surface area formula, the surface area is

$$\int_0^3 2\pi x^2\sqrt{1+4x^2}\,dx.$$

□

Example 8.5.4 Gabriel's Horn. Find the volume of the solid of revolution obtained by rotating the graph of $y = \frac{1}{x}$ about the x-axis from $x = 1$ onwards. Estimate the surface area of this solid of revolution.

Solution. The volume is given by

$$\int_1^\infty \frac{\pi}{x^2}\,dx = \pi.$$

The integral that gives the surface area is

$$\begin{aligned}\int_1^\infty \frac{2\pi}{x}\sqrt{1+\frac{1}{x^2}}\,dx &> \int_1^\infty \frac{2\pi}{x}\,dx \\ &= \infty\end{aligned}$$

Oddly enough, the volume of this region is finite, but the surface area is *infinite*. A silly way to describe this situation: you can fill this region with a finite amount of paint, but it would take an infinite amount to paint the surface. □

Example 8.5.5 Find the surface area of the solid generated by rotating the graph of $y = x^3$ from $x = 1$ to $x = 2$ about the x-axis.

Solution. Since $f(x) = x^3$, then $f'(x) = 3x^2$ and the surface area of the solid is

$$\int_1^2 2\pi x^3\sqrt{1+9x^4}, dx.$$

□

8.6 Applications to Physics and Engineering

The primary use of integrals in physics and engineering is to measure the accumulation of continuously varying quantities. The running theme in this section:

1. Measure the accumulation of some quantity by approximating it as a (finite) sum of ”small” values.

2. Improve the approximation by taking more and more values.

3. Introduce a limit to make the approximation exact, resulting in an integral.

Work

We introduced a formula for work in Section 7.4, and now we'll actually derive it. First, recall that **work** (very roughly) represents the net effect of a **force** F acting on a mass over some displacement d. If the force is constant (in both magnitude and direction), then we can simply define the work W accomplished by a force F as a particle is displaced by d units by the formula

$$W = Fd.$$

So with this definition, work is proportional to the magnitude of the force F as well as the displacement d over which the force acts. If F is measured in units of $\frac{\mathrm{m}}{\mathrm{s}^2}$ or more simply N, and d is measured in units of m the W takes on units of $\mathrm{N} \cdot \mathrm{m}$ or just J.

Although there are important forces that are reasonably approximated by constants, such as the force due to gravity, this isn't reasonable for every force of interest. For example, consider the force due to drag against an object moving through the atmosphere, which should increase as velocity increases. Another basic example is the **Hooke's Law**, which states that the force a spring exerts on an attached mass is proportional to the displacement of the mass from the spring's equilibrium position. For these non-constant forces, our old formula doesn't work so well. To be precise, how can we measure the work done by a non-constant force $F(x)$ on a particle as it moves from $x = a$ to $x = b$?

As you might guess from this section's running theme, we'll start by *approximating* the work done by measuring work in small displacements as the particle moves from a to b. Over small enough intervals, even variable forces will look almost constant, which makes $W = Fd$ a reasonable estimate of the work done (on this small interval!). So let's start by breaking up $[a, b]$ into n smaller subintervals

$$[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n],$$

where $a = x_0 < x_1 < \ldots < x_n = b$, and let $\Delta x_i = x_i - x_{i-1}$ be the width of each subinterval. Now choose sample points x_i^* from each subinterval. Once again, if Δx_i is small enough, then the work done by the force over $[x_{i-1}, x_i]$ should be approximately

$$F(x_i^*)\Delta x_i.$$

This is the approximate work for one subinterval, so adding in the approximations from the other subintervals also provides an estimate of the total work W done by $F(x)$ over $[a, b]$. That is,

$$W \approx F(x_1^*)\Delta x_1 + \cdots + F(x_n^*)\Delta x_n = \sum_{i=1}^{n} F(x_i^*)\Delta x_i.$$

As $\Delta x_i \to 0$, the approximation becomes exact. Therefore

$$W = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} F(x_i^*)\Delta x_i = \int_a^b F(x)\,dx. \tag{II.4}$$

Example 8.6.1 A mass attached to a spring. A mass attached to a spring is acted upon by a **spring force** given by $F(t) = -2x$, where x is the displacement of the mass from the equilibrium position of the spring. Suppose that the mass is held 1 m to the right of the spring's equilibrium position (so that the spring will try to pull the mass back towards its equilibrium which we'll say occurs at $x = 0$). What is the work done by the spring on the mass in moving the mass from its current position back to equilibrium?

Solution. Since equilibrium occurs at $x = 0$, the work done should be

$$W = \int_1^0 (-2x)\, dx = \int_0^1 2x\, dx = 1.$$

□

In Example 8.6.1, the reason the work done was positive was because the force and the direction the mass was moving tended to align. Once the mass moves past equilibrium and to the left, the force will begin acting against the motion of the mass resulting in negative work over that particular interval. The spring force itself is an example of a **restoring force**, since it's always trying to pull the mass back to equilibrium. To highlight this, we often write the spring force as $F_S = -kx$ for some positive k.

Moments and Centers of Mass

Consider a system of two point masses m_1, m_2 connected by a rod of negligible mass (so we can assume that the mass of the system is just $m_1 + m_2$). These masses are located at the points x_1, x_2 on the x-axis. The rod is then placed upon a fulcrum at the point $x = c$, as shown below:

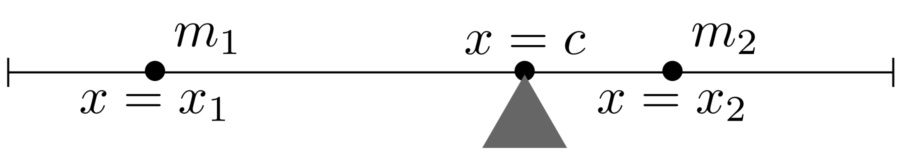


Figure 8.6.2 Rod on a fulcrum.

Our goal is to find the center of mass of the rod.

Suppose that the first mass is d_1 units away from the fulcrum, and that the second mass is d_2 units away. Archimedes' Law of the Lever states that the rod will balance if $m_1 d_1 = m_2 d_2$. Since $d_1 = c - x_1$ and $d_2 = x_2 - c$ (since d_1, d_2 are positive!), we can rewrite this as

$$\begin{gathered} m_1 d_1 = m_2 d_2 \\ \Rightarrow \\ m_1(c - x_1) = m_2(x_2 - c) \\ \Rightarrow \\ m_1 c - m_1 x_1 = m_2 x_2 - m_2 c \\ \Rightarrow \\ m_1 c + m_2 c = m_1 x_1 + m_2 x_2 \end{gathered}$$

which simplifies down to

$$c = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}.$$

Hence c is the location of the center of mass of the rod if this equation is true!

More generally, if n point masses m_i at position $x = x_i$ are connected by a rod of negligible mass, then the center of mass $\overline{x}$ of the resulting system is given by

$$\overline{x} = \frac{m_1x_1 + m_2x_2 + \cdots + m_nx_n}{m_1 + m_2 + \cdots + m_n} = \frac{\sum_{i=1}^n m_ix_i}{m},$$

where $m = \sum_{j=1}^n m_j$ is the total mass of the system and

$$M = \sum_{i=1}^n m_ix_i$$

is called the **moment of the system about the origin**.

Example 8.6.3 Consider the following system of three point masses m_1, m_2, m_3 at positions x_1, x_2, x_3 given as follows:

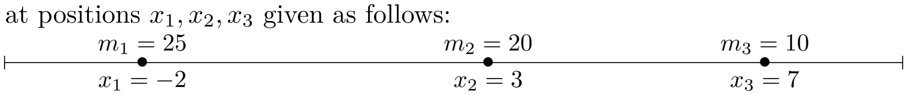


Figure 8.6.4 Three point masses.

Find the following:

1. The moment of the system about the origin.
2. The center of mass.

Solution.

1. The moment of the system about the origin is defined to be

$$\begin{aligned} M &= \sum_{i=1}^3 m_ix_i \\ &= 25 \cdot (-2) + 20 \cdot (3) + 10 \cdot (7) \\ &= 80. \end{aligned}$$

So the moment is $M = 80$.

2. The center of mass is just the moment divided by the total mass m of the system:

$$\begin{aligned} \overline{x} &= \frac{M}{m} \\ &= \frac{80}{25 + 20 + 10} \\ &= \frac{80}{55} \\ &= \frac{16}{11}. \end{aligned}$$

So the center of mass of the rod is at the point $\overline{x} = \frac{16}{11}$.

□

We now move from one dimension to two. Consider a system of point masses $m_1, \ldots, m_n$ at the points $(x_1, y_1), \ldots, (x_n, y_n)$, and suppose that these masses are connected by a thin plate of negligible mass. The **moment of the system about the *y*-axis** is the number M_y given by

$$M_y = \sum_{i=1}^n m_ix_i.$$

This can be thought of as tendency of the system to rotate *about* the y-axis. Increasing the mass or the distance from the y-axis increases this tendency.

The **moment of the system about the x-axis** is the number M_x given by

$$M_x = \sum_{i=1}^{n} m_i y_i.$$

Just as in the one-dimensional case, the center of mass for this system will be determined by the total mass and the above moments. In particular, the center of mass of this system is located at the point $(\overline{x}, \overline{y})$ with coordinates

$$\overline{x} = \frac{M_y}{m} \quad \text{and} \quad \overline{y} = \frac{M_x}{m},$$

where m is the total mass $m_1 + \cdots + m_n = \sum_{i=1}^{n} m_i$.

Example 8.6.5 Four point masses $m_1 = 6, m_2 = 5, m_3 = 1, m_4 = 4$ are located in the plane at the points $(1, -2), (3, 4), (-3, -7)$ and $(6, -1)$.

1. Find the moments of the system about the y-axis and the x-axis.
2. Find the center of mass of this system.

Solution.

1. The moment about the y-axis is

$$\begin{aligned} M_y &= \sum_{i=1}^{4} m_i x_i \\ &= 6 \cdot 1 + 5 \cdot 3 + 1 \cdot (-3) + 4 \cdot 6 \\ &= 42. \end{aligned}$$

   Similarly, the moment about the x-axis is given by

$$\begin{aligned} M_x &= \sum_{i=1}^{4} m_i y_i \\ &= 6 \cdot (-2) + 5 \cdot 4 + 1 \cdot (-7) + 4 \cdot (-1) \\ &= -3. \end{aligned}$$

2. The center of mass is the point $(\overline{x}, \overline{y})$ where

$$\overline{x} = \frac{M_y}{m} = \frac{42}{16} \quad \text{and} \quad \overline{y} = \frac{M_x}{m} = \frac{-3}{16}$$

   and m is the total mass of the system.

$\square$

} We now consider a more complicated situation than the above. Suppose that we have a thin, flat plate (or *lamina*) R of uniform density in the plane. Our goal now is to find the center of mass, or **centroid**, of the plate. To do so, we will use the following fact: *the centroid of a rectangular lamina of uniform density is just the center of the rectangle.* Now consider a lamina R with uniform density ρ in the xy-plane and bounded between $x = a, x = b$ and the functions $f(x)$ and $g(x)$:

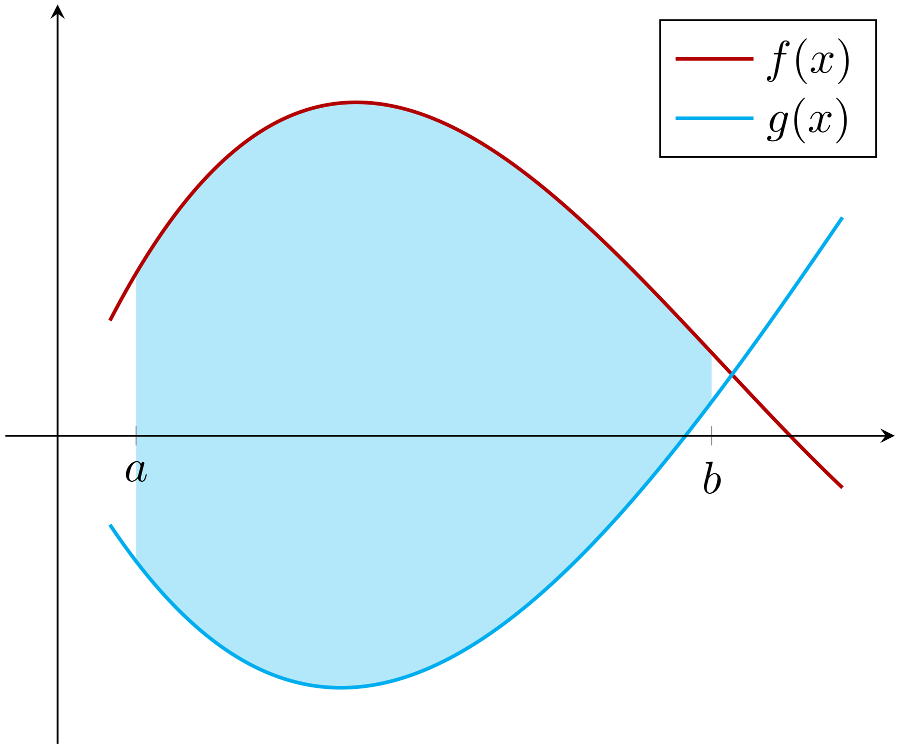


Figure 8.6.6 A lamina in $\mathbb{R}^2$.

To find the centroid of this lamina we will break it up into (vertical) approximating rectangles $R_0, R_1, \ldots, R_n$, each of which has width Δx and centroid (x_i, y_i):

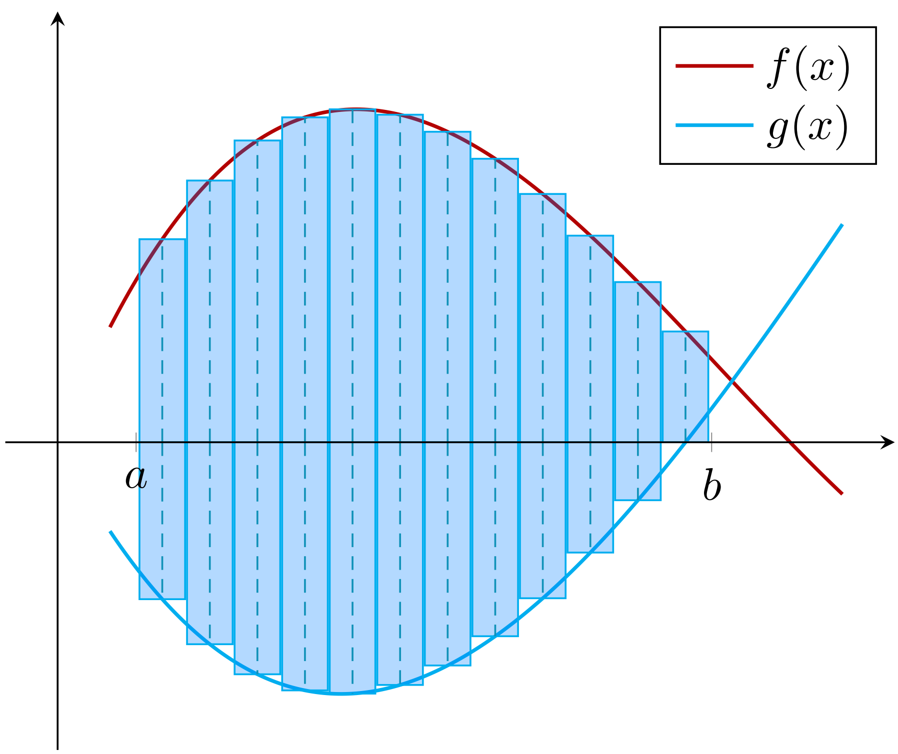


Figure 8.6.7 Breaking the lamina into subrectangles.

Then the lamina may be approximated by a system of point masses $m_0, \ldots, m_n$, with m_i located at the centroid rectangle R_i and having the mass of the corresponding rectangle. To find the center of mass of the lamina, we will approximate it with the center of mass of the system of point masses.

To do so, we will need:

$$m = \sum_{i=0}^{n} m_i$$
$$M_y = \sum_{i=0}^{n} m_i x_i$$
$$M_x = \sum_{i=0}^{n} m_i y_i$$

Consider a single rectangle R_i:

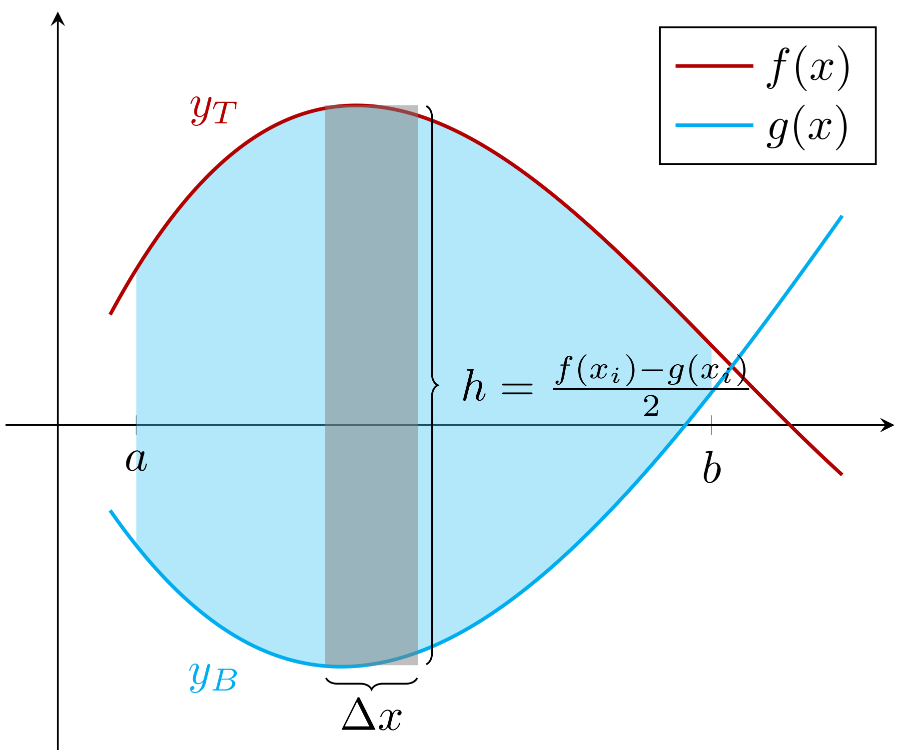


Figure 8.6.8 A single rectangle.

Then we have

$$m_i = (\text{density})(\text{area of } R_i) = \rho \frac{f(x_i) - g(x_i)}{2} \Delta x$$
$$y_i = \frac{f(x_i) + g(x_i)}{2}.$$

We can now fill in the following table:

Table 8.6.9 Moments and center of mass

	System of point masses	Lamina
m	$\sum_{i=0}^{n} m_i = \sum_{i=0}^{n} \rho[f(x_i) - g(x_i)]\Delta x$	$\rho \int_a^b [f(x) - g(x)]\, dx$
M_y	$\sum_{i=0}^{n} m_i x_i = \sum_{i=0}^{n} \rho x_i [f(x_i) - g(x_i)]\Delta x$	$\rho \int_a^b x[f(x) - g(x)]\, dx$
M_x	$\sum_{i=0}^{n} m_i y_i = \sum_{i=0}^{n} \rho[f(x_i) - g(x_i)]\frac{f(x_i)+g(x_i)}{2}\Delta x$	$\frac{\rho}{2} \int_a^b [f(x)^2 - g(x)^2]\, dx$

The center of mass for this lamina is then the point $(\overline{x}, \overline{y})$ where

$$\overline{x} = \frac{M_y}{m} = \frac{\int_a^b x[f(x) - g(x)]\, dx}{\int_a^b [f(x) - g(x)]\, dx}$$

$$\overline{y} = \frac{M_x}{m} = \frac{1}{2} \frac{\int_a^b [f(x)^2 - g(x)^2]\, dx}{\int_a^b [f(x) - g(x)]\, dx}$$

8.7 Differential Equations

Basic Concepts

Definition 8.7.1 Differential Equations. A **differential equation** is an equation relating some function with its derivatives. A differential equation that involves a function of only one independent variable is called an **ordinary differential equation**, or ODE. A differential equation that involves a function of more than one independent variable (which you see a lot of in Calculus 3) is called a **partial differential equation**, or PDE. The **order** of a differential equation is the highest derivative that appears in the equation. ♢

Examples of ODEs:

- $\dfrac{d^2y}{dx^2} + y = 0$; this is a second order ODE relating the unknown function y with its second derivative.
- $5t^2x^{'''} - e^x = 3t$; this is a seventh order ODE involving the derivatives of the unknown function x. Note that in this ODE t is the independent variable whereas x serves as the dependent variable.

Differential equations serve as useful *mathematical models* for quantities that change over time. In particular, Newton's Second Law leads to a number of significant differential equations.

Example 8.7.2 The Spring Equation. Consider an object of mass m attached to a spring, and suppose that the spring force F_S is the only force acting on the mass. Assuming that Hooke's Law describes the spring force, find a differential equation modeling the motion of the mass.

Solution. Recall that Hooke's Law says that $F_S = -kx$, where k is a positive constant and x is the displacement of the mass from equilibrium. If this is the only force acting on the mass, then by Newton's Second Law

$$F_{\text{net}} = ma \Rightarrow F_S = mx''.$$

Therefore the displacement x satisfies the differential equation

$$mx'' + kx = 0.$$

□

Definition 8.7.3 Solution of a Differential Equation. A function is a **solution** of a differential equation if it satisfies the differential equation. ◊

It is straightforward to check if a function is a solution of some given differential equation, but *finding* solutions involves a bit more work.

Example 8.7.4 Verifying solutions. Is $y = 5e^{x^2}$ a solution of the ODE $y' = 2xy$?

Solution. At this point we don't know how to solve differential equations, but that doesn't mean we can't *check* solutions of differential equations. To do so, we just plug $5e^{x^2}$ wherever y shows up in the ODE and see if the resulting equation is true. So we have

$$\begin{aligned} y' &= 2xy \\ (5e^{x^2})' &= 2x(5e^{x^2}) \text{ after substituting } y = 5e^{x^2} \\ 10xe^{x^2} &= 10xe^{x^2} \text{ after simplifying} \end{aligned}$$

This is a true statement, so $y = 5e^{x^2}$ satisfies the ODE. Hence $5e^{x^2}$ is a solution of the ODE. □

In Example 8.7.4, $y = 5e^{x^2}$ is not the only solution of $y' = 2xy$. You can check that $y = 3e^{x^2}$ and $y = -10e^{x^2}$ are also solutions. In fact, *any* function of the form Ce^{x^2} where C is a constant is a solution of $y' = 2xy$.

To specify a unique solution to a differential equation, we need to add another condition known as an *initial condition* or *initial value*, often of the form $y(x_0) = y_0$.

Definition 8.7.5 An ODE together with an initial condition is known as an **initial value problem**, or IVP. ◊

Separable Equations

Many ODEs are difficult, if not impossible, to solve exactly. The simplest ODEs to solve are the first-order ODEs of the form $\frac{dy}{dx} = f(x)$. The Fundamental Theorem of Calculus guarantees that the solution y is given by $y = \int f(x)\,dx$.

Another type of ODE that is relatively straightforward to solve is the **separable ODE**, which is a first-order ODE that can be written in the form

$$\frac{dy}{dx} = f(x)g(y).$$

These ODEs can be solved by integration as well, but only after some rearranging.

Example 8.7.6 Solving a separable ODE. Solve the IVP given by $y' = x + 4xy^2, y(0) = 1$.

Solution. The first step to solving this IVP is to solve the ODE $y' = x+4xy^2$. It may not look like it at first, but this ODE is separable since we can rewrite it as $y' = x(1 + 4y^2)$. To solve this ODE, we need to move the y terms to the left hand side of the equation and the x terms to the right hand side. We'll abuse notation a little bit to do so by rewriting $y' = \frac{dy}{dx}$ and treating $\frac{dy}{dx}$ as a fraction, but it won't get us into too much trouble here:

$$\begin{aligned} \frac{dy}{dx} = x(1 + 4y^2) &\Rightarrow \frac{dy}{1 + 4y^2} = x\,dx \\ &\Rightarrow \int \frac{dy}{1 + 4y^2} = \int x\,dx \end{aligned}$$

$$\Rightarrow \frac{1}{2}\tan^{-1}(2y) = \frac{1}{2}x^2 + C$$
$$\Rightarrow \tan^{-1}(2y) = x^2 + C_1$$

At this step we can either leave the solution as is (in **implicit form**) or solve for y to get an **explicit form**. We'll leave this in implicit form and then plug in the initial condition to get

$$\tan^{-1}(2) = C_1.$$

So the implicit solution of this IVP is given by

$$\tan^{-1}(2y) = x^2 + \tan^{-1}(2).$$

□

In Example 8.7.6, we can also find an *explicit* form of the solution by solving for y:

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

Population Equations

Suppose we're monitoring the population of some species, and let's denote the population at time t by $P(t)$. An obvious question to consider is how that population will change over time. Mathematically, this means we want to obtain information on $\frac{dP}{dt}$ and then use it to estimate $P(t)$.

A simple model for $\frac{dP}{dt}$ is to assume it depends only on the birth rate β and death rate δ of the species in question. Then we can write

$$\frac{dP}{dt} = (\beta - \delta)P. \tag{II.5}$$

If we assume that β, δ are constants, then this equation is separable and we can solve it to obtain

$$P(t) = P_0 e^{(\beta-\delta)t},$$

where P_0 represents the "initial population", or population at time $t = 0$. We call (II.5) the **natural growth equation**.

The natural growth equation is simple, but it's probably too simple to be useful expect in certain scenarios (such as measuring half-life). To get a more flexible model, we can generalize (II.5) by assuming that the birth and death rates are actually functions of time. This gives us the **general population equation**.

Definition 8.7.7 General Population Equation. The general population equation for a population $P(t)$ is given by

$$\frac{dP}{dt} = (\beta(t) - \delta(t))P.$$

♢

Example 8.7.8 Population Explosion. A population has 100 members at time $t = 0$ years with a death rate of $.25P$ and a birth rate of $.5P$, where $P(t)$ denotes the population after t years. Find $P(t)$ and determine if this is a reasonable population model.

Solution. If we assume that the population obeys the general growth equation, then we get

$$P' = .25P^2, P(0) = 100.$$

This ODE is separable, and we can therefore solve it to get

$$P(t) = \frac{100}{1-25t}.$$

So we have a solution, and it can be shown that the solution is unique. But if you stare at this for a bit, you might see that it has a divide-by-zero problem. In particular,

$$\lim_{t \to (\frac{1}{25})^+} P(t) = \infty.$$

In other words, the population becomes infinite in about two weeks! □

The Logistic Equation

Example 8.7.8 shows that we need to be more careful with our assumptions on population growth. One relatively simple assumption we can make is to assume that the birth rate $\beta(t)$ decreases as population P increases. This makes sense in the physical world as well: as population increases, existing and finite resources (such as food) must be shared between more and more members of the population. Since there's less to go around, we should expect growth to slow down. In particular, let's assume that

$$\beta(t) = \beta_0 - \beta_1 P$$
$$\delta(t) = \delta_0$$

where β_0, β_1 and δ_0 are all positive constants. Then the population equation for this scenario becomes

$$\frac{dP}{dt} = (\beta_0 - \beta_1 P - \delta_0)P.$$

With a little algebra, we get the **logistic equation**:

$$\frac{dP}{dt} = kP(M-P)$$

for constants k and M. This equation is separable, and can be solved using partial fractions to obtain

$$P(t) = \frac{MP_0}{P_0 + (M-P_0)e^{-kMt}},$$

where $P_0 = P(0)$. In order to verify the reasonableness of our logistic model, let's see what happens to the population as time increases.

Example 8.7.9 Long-Term Behavior of Logistic Growth. What is the long-term population of a species that grows according to the logistic equation $\dfrac{dP}{dt} = kP(M-P)$?

Solution. Using the fact that

$$P(t) = \frac{MP_0}{P_0 + (M-P_0)e^{-kMt}},$$

we have

$$\lim_{t\to\infty} P(t) = M.$$

So the population should eventually level out at M. $\square$

In the logistic equation $P' = kP(M - P)$, the value M is the **carrying capacity**, and denotes the maximum sustainable population according to the model.

Example 8.7.10 Population Growth in the USA. In millions, the population of the USA in 1990 was 250 and was growing at a rate of 3.1 per year. In 2012, the population was 314 and was growing at a rate of 2.3 per year. Assuming that the population of the USA grows logistically, estimate the population of the USA in 2017 and compare it to the current estimate of 325.7.

Solution. Let $P(t)$ denote the population of the USA (in millions), where t is the number of years after 1990. Then

$$\frac{dP}{dt} = kP(M - P)$$

and

$$P(t) = \frac{MP_0}{P_0 + (M - P_0)e^{-kMt}}.$$

So we need to find k and M.

When $t = 0$, we have $P' = 3.1$ and $P = 250$. Similarly, when $t = 22$ we have $P' = 2.3$ and $P = 314$. Therefore

$$\begin{aligned} 3.1 &= 250k(M - 250) \\ 2.3 &= 314k(M - 314) \end{aligned}$$

Solving this system gives us $k \approx .00008$ and $M \approx 406.4$. Hence

$$P(t) = \frac{101600}{250 + 156.4e^{-.03t}}.$$

This model estimates the population in 2017 to be

$$P(27) = 317.9,$$

which is about a 2% error. Note also that this model predicts the carrying capacity of the USA to be 406.4. $\square$

Stability of Solutions

The logistic equation

$$\frac{dP}{dt} = kP(M - P)$$

is a particularly nice separable ODE since the right hand side depends only on the unknown function P. So we can write $P' = f(P)$, where $f(P) = kP(M - P)$. ODEs like this (where the independent variable does not appear explicitly) are called **autonomous ODEs**.

Autonomous ODEs like $\dfrac{dx}{dt} = f(x)$ are useful because the behavior of their solutions can be determined *qualitatively*, without actually solving the ODE. This is done by looking for the constant solutions of the ODE, that is, solutions of the form $x = c$. For any such solution, we must have $f(c) = 0$ as well. These solutions (i.e., the solutions of $f(x) = 0$) are called the **critical points** or **equilibrium solutions** of the ODE. These solutions completely determine the long-term behavior of *every other solution*.

Example 8.7.11 Finding Equilibrium Solutions. Find the equilibrium solutions of $\dfrac{dx}{dt} = -x^2 + 7x - 10$.

Solution. We need to solve the equation $-x^2 + 7x - 10 = 0$. Thankfully, we can factor this to get $(2 - x)(x - 5) = 0$, and so the equilibrium solutions are $x = 2, 5$. □

Definition 8.7.12 Stability of Solutions. A critical point is **stable** if solutions that start "near" the point stay near it. A critical point is **unstable** if solutions that start "near" the point can diverge away from it. ♢

Example 8.7.13 Determining the Stability of Solutions. What are the stable critical points of $\dfrac{dx}{dt} = -x^2 + 7x - 10$?

Solution. We already know that the critical points are $x = 2, 5$. We can determine their stability by making use of a **phase diagram**, which is essentially a sign chart for $f(x) = -x^2 + 7x - 10$:

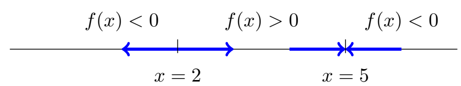


Figure 8.7.14 The phase diagram for $x' = f(x)$.

This shows us that solutions that begin near $x = 2$ tend to move away from $x = 2$, which solutions near $x = 5$ tend to move towards $x = 5$. So $x = 2$ is unstable and $x = 5$ is stable. □

Example 8.7.15 Determining a Sustainable Population. Consider a population of fish that obeys the logistic equation

$$\frac{dP}{dt} = 2P(30 - P)$$

where $P(t)$ is the population of fish (in thousands) after t years. Suppose that the population is also harvested at some rate h (in thousands per year). What is the maximum sustainable rate of harvesting?

Solution. To account for the harvesting, we need to modify the ODE:

$$\frac{dP}{dt} = 2P(30 - P) - h.$$

The harvesting will be sustainable as long as the population does not become extinct. To determine this long term behavior, we'll find the critical points and set up a phase diagram.

The critical points are given by

$$P = 15 \pm \sqrt{3600 - 8h}$$

by the quadratic formula. We now have three cases to consider: $3600 - 8h < 0, 3600 - 8h = 0, 3600 - 8h > 0$. In terms of h, these reduce to $h < 450, h = 450, h > 450$.

1. In the first case, if $h < 450$ then we have two positive, real critical points:

$$0 < c_1 < 15 < c_2 < 75.$$

The phase diagram for this situation is

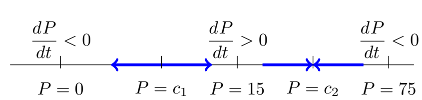


Figure 8.7.16 Phase diagram for $h < 450$.

So we see that c_1 is unstable while c_2 is stable. In particular, as long as $P \geq c_1 = 15 - \sqrt{3600 - 8h}$, then the rate of harvesting is sustainable.

2. Now assume that $h = 450$. Then we have only one equilibrium solution: $c = 15$. The corresponding phase diagram is

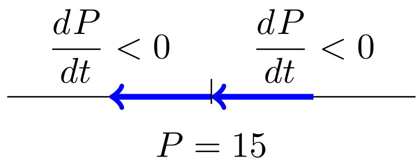


Figure 8.7.17 Phase diagram for $h = 450$.

We interpret the phase diagram as follows: if P is less than 15,000then the population will collapse to extinction. Otherwise, the population will stabilize at $15,000$. This type of critical point is often called **semi-stable.**

3. Finally, consider the case $h > 450$. Then we have no (real) critical points. Since imaginary populations don't make sense in this model, there is no sustainable population. No matter how large the initial population, it will eventually go extinct if harvested at a rate greater than 450.

By the above, the largest sustainable harvesting rate is $h = 450$, as long as $P_0 \geq 15$. □

Linear Stability Analysis

Given the autonomous ODE $\frac{dx}{dt} = f(x)$, we saw above that we can qualify the behavior of equilibrium solutions by setting up a phase diagram. We can go a step further and actually qualify the growth of solutions that are "near" equilibrium solutions. In particular, we have the following theorem.

Theorem 8.7.18 Linear Stability Analysis. *Suppose* $\frac{dx}{dt} = f(x)$ *where* $f(x)$ *is continuously differentiable, and let* x^* *denote a critical point/equilibrium solution of the ODE. If* $f'(x^*) < 0$*, then* x^* *is stable and solutions near* x^* *will move exponentially towards* x^**. If* $f'(x^*) > 0$*, then* x^* *is unstable and solutions near* x^* *will move exponentially away from* x^**. If* $f'(x^*) = 0$*, then more advanced methods are required.*

Example 8.7.19 Classifying the Critical Points of the Logistic Equation. Classify the critical points of the logistic equation as stable or unstable.

Solution. Recall that the logistic equation is given by $P' = kP(M - P) = f(P)$ for (we'll assume) positive constants k, M. From here, we clearly see that the critical points are $P = 0$ and $P = M$ (which makes sense from a population standpoint!). We could set up a phase diagram to determine stability, but we'll use Theorem 8.7.18 instead.

Since $f'(P) = k(M - P) - kP$, we see that

$$f'(0) = kM > 0$$
$$f'(M) = -kM < 0$$

Hence $P = 0$ is unstable, while $P = M$ is stable. □

Chapter 9

Series

In physics, the motion of a simple pendulum can be modeled using the second-order differential equation

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0$$

where g is the acceleration due to gravity,l is the length of the pendulum rod and θ is the angular displacement of the pendulum. Unfortunately, it turns out that this ODE is impossible to solve exactly in terms of "elementary functions". The sine term is, surprisingly, too complicated to handle by standard methods for solving ODEs.

However, we can *replace* the sine term with a reasonable approximation, namely the linear approximations from Section 3.7. The linear approximation $L(\theta)$ to $\sin\theta$ at $\theta = 0$ is given by

$$L(\theta) = \sin 0 + \frac{d}{d\theta}\sin\theta\Big|_{\theta=0}(\theta - 0),$$

or just $L(\theta) = \theta$ (see Figure 9.0.1). This replaces the pendulum ODE with the approximation

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\theta = 0,$$

which is much simpler and can be easily solved (using methods from differential equations) to get

$$\theta(t) = c_1\cos\sqrt{\frac{g}{l}}t + c_2\sin\sqrt{\frac{g}{l}}t.$$

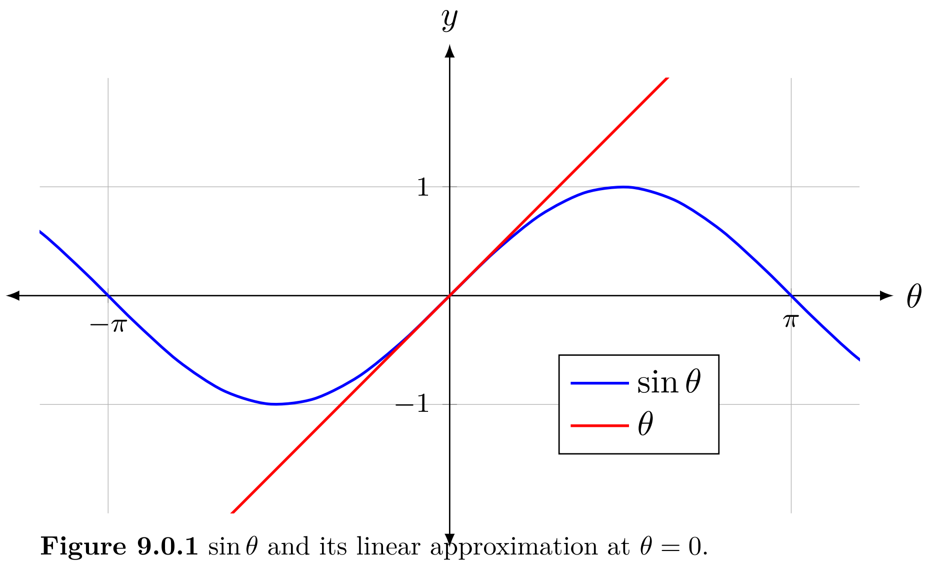


Figure 9.0.1 $\sin\theta$ and its linear approximation at $\theta = 0$.

As we can see from Figure 9.0.1, the linear approximation serves as a reasonable estimate of $\sin\theta$ if θ is small, but loses accuracy quickly as θ moves away from 0. A reasonable question then is how we can improve the accuracy of our approximation. One way to do so is to find a *quadratic* approximation to $\sin\theta$, the idea being that allowing higher powers of θ in our approximation will remove some of the error.

To find the "best" quadratic approximation to $f(\theta) = \sin\theta$ at $\theta = 0$, we want to find the quadratic function $g(\theta) = a\theta^2 + b\theta + c$ that satisfies the following:

- $f(0) = g(0)$.
- $f'(0) = g'(0)$.
- $f''(0) = g''(0)$.

These three conditions give

- $0 = c$
- $1 = b$
- $0 = 2a$

So the best quadratic approximation to $\sin\theta$ (in the sense of matching derivatives up to the second order) is given by

$$g(\theta) = 0\theta^2 + \theta + 0 = \theta.$$

This is the same as the linear approximation, so we'll go one more step up to get the best cubic approximation to $f(\theta) = \sin\theta$, i.e., the best approximation of the form $h(\theta) = a\theta^3 + b\theta^2 + c\theta + d$ in the sense that

- $f(0) = h(0)$
- $f'(0) = h'(0)$
- $f''(0)$
- $f'''(0) = g'''(0)$

This simplifies to

- $0 = d$
- $1 = c$
- $0 = 2b$
- $-1 = 6a$

So the best cubic approximation to $\sin\theta$ is given by

$$h(\theta) = -\frac{1}{6}\theta^3 + 0\theta^2 + \theta - 0 = \theta - \frac{1}{6}\theta^3.$$

Plotting this against $\sin\theta$ and θ as in Figure 9.0.1 gives Figure 9.0.2:

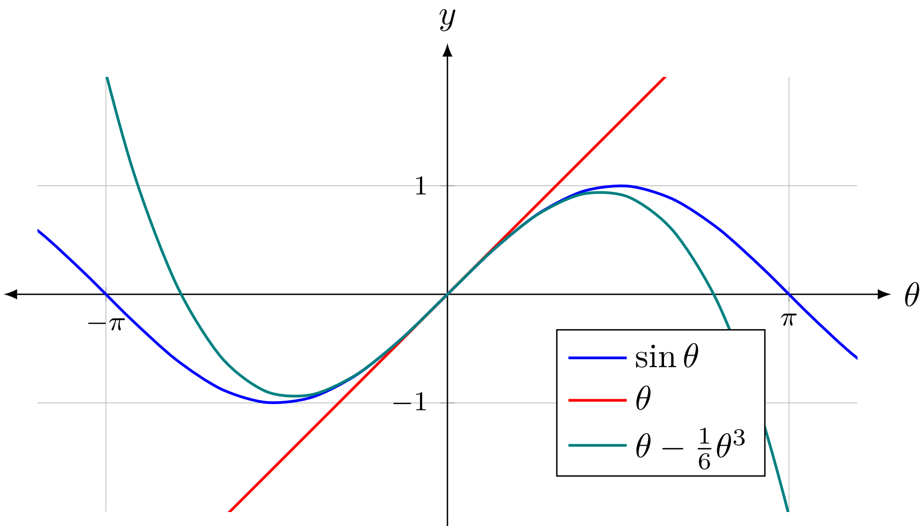


Figure 9.0.2 $\sin\theta$ and its linear and cubic approximations at $\theta = 0$.

It turns out that this process can be continued indefinitely, giving us

$$\sin x = \sum_{n=0}^{\infty} a_n x^n.$$

The expression on the right is known as a **power series**, and provides many advantages for performing calculus with the sine function. The list of coefficients $\{a_n\}_{n=0}^{\infty}$ is known as a **sequence**, and completely determines the properties of sine.

9.1 Sequences

In this chapter we'll be using infinite sums of the form $\sum_{k=0}^{\infty} a_k x^k$ to represent functions and compute integrals. In order to make sense of these series, we need to introduce the concept of a sequence and the limit of a sequence.

A **sequence** is a list of numbers:

$$(a_0, a_1, a_2, \ldots),$$

often written $\{a_n\}_{n=0}^{\infty}$.

We will often take $n = 0$ as our starting index, but not always.

We call a_n the $\mathbf{n^{th}}$ term of the sequence, and n itself the **index**. We can view n as denoting the position of the number a_n within the sequence.

Example 9.1.1 Finding a Formula for a Sequence. Given the sequence $(a_n)_{n=0}^{\infty} = (1, 2, 4, 8, \ldots)$, make a reasonable guess of the value of a_4 and the general formula for a_n. □

Sequences are usually specified in one of two ways: as an **explicit formula** such as

$$a_n = \frac{1}{1-2n},$$

or recursively by means of a **recurrence relation**, such as

$$F_n = F_{n-1} + F_{n-2}, \quad f_0 = f_1 = 1.$$

Note that for recurrence relations, we need to specify "base cases".

Example 9.1.2 An Alternating Sequence. Find the first few terms of the sequence $\{\cos(n\pi)\}_{n=2}^{\infty}$.

Solution. This sequence simplifies down to $\{1, -1, 1, -1, 1, \ldots\}$. □

Sequences have limits just as functions do.

Definition 9.1.3 Limit of a Sequence. A sequence $\{a_n\}$ has limit L, denoted

$$\lim_{n\to\infty} a_n = L \quad \text{or} \quad a_n \to L,$$

if a_n gets arbitrarily close to L as n increases. If a sequence has a limit, we say the sequence is **convergent** and **converges**. Otherwise, we say the sequence is **divergent** and **diverges**. ♢

Graphically, we can say that a sequence a_n has a limit L if the points (n, a_n) become arbitrarily close to the line $y = L$:

```
# plot of cos(n)/(n^1.5)
# limit is L = 0
points([(n, cos(n)/n^(3/2)) for n in range(1, 100)],
    legend_label = r'$a_n = \frac{\cos(n)}{n}$')
```

If you run the above code cell, you get some pretty convincing evidence that $\frac{\cos(n)}{n^{3/2}} \to 0$ as $n \to \infty$.

One of the most important sequential limits is the following:

$$\frac{1}{n^p} \to 0 \quad \text{as} \quad n \to \infty$$

if $p > 0$. Many limits involving sequences with terms that are rational functions of n can be reduced to this form when finding limits.

Example 9.1.4 Find the limit of the sequence

$$a_n = \frac{3n - n^3}{5n^3 + 2n^2}.$$

Solution. We can try dividing the numerator and denominator by the highest power of n that appears: n^3. This gives

$$a_n = \frac{3n^{-2} - 1}{5 + 2n^{-1}}.$$

Now we can take the limit as $n \to \infty$ to get $a_n \to \frac{-1}{5}$. □

We can also apply Calculus 1 limits to sequences by using the following theorem.

Theorem 9.1.5 Sequential and Functional Limits. *Let $f(x)$ be a function and suppose that*

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

Then

$$\lim_{n\to\infty} f(n) = L$$

also.

One immediate advantage of Theorem 9.1.5 is that L'Hospital's Rule from Section 4.6 applies to sequential limits as well, *as long as the sequence consists of values from a differentiable function.*

Example 9.1.6 Finding a Sequential Limit Using L'Hospital's Rule. Let $a_n = n^2 e^{-n}$. Find $\lim_{n\to\infty} a_n$.

Solution. First, note that $a_n = f(n)$ where

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

Therefore

$$\begin{aligned} \lim_{n\to\infty} a_n &= \lim_{x\to\infty} \frac{x^2}{e^x} \\ &= \lim_{x\to\infty} \frac{2x}{e^x} \\ &= \lim_{x\to\infty} \frac{2}{e^x} \\ &= 0. \end{aligned}$$

□

Example 9.1.7 Geometric Sequences. A **geometric sequence** is a sequence of the form

$$a_n = ar^n.$$

Find $\lim_{n\to\infty} a_n$.

Solution. This limit depends on whether or not r is in $(-1, 1]$. If $|r| < 1$ then $r^n \to 0$. If $r = 1$ then $r^n = 1$ for all n. Finally, if r is outside of this interval, then r^n diverges. Therefore

$$a_n \to \begin{cases} 0 & \text{if } -1 < r < 1 \\ a & \text{if } r = 1 \end{cases}$$

and diverges otherwise. □

May decimals can be represented using geometric sequences.

Example 9.1.8 .9 **Repeating.** Determine the limit of the sequence

$$\{.9, .99, .999, .9999, .99999, \ldots\}.$$

Solution. It looks like the terms of the sequence are approaching 1, and we can verify this using a geometric sequence. We can write this sequence as

$$a_n = 1 - (.1)^n \quad \text{for} \quad n \geq 1.$$

So the limit of the sequence is

$$\lim_{n\to\infty} a_n = 1.$$

Note that this suggests the (true!) statement that $.9999\ldots = 1$. □

Definition 9.1.9 Infinite Limits. Let a_n be a sequence. If the terms of a_n grow without bound as n increases, we say that $a_n \to \infty$. If the terms of a_n decrease without bound as n increases, we say that $a_n \to -\infty$. $\Diamond$

Example 9.1.10 Limit of the Fibonacci Sequence. Let F_n denote the n^{th} term of the Fibonacci sequence. Determine $\lim_{n\to\infty} F_n$. Estimate $\lim_{n\to\infty} \frac{F_n}{F_{n-1}}$.

Solution. One approach to estimate the limit is to graph the ratio $\frac{F_n}{F_{n-1}}$ to see if it approaches a limiting value. A computer system can handle this easily. The values of $\frac{F_n}{F_{n-1}}$ appear to settle in quickly around $L \approx 1.6$.

The actual limiting value is $\varphi = \frac{1+\sqrt{5}}{2}$, the **golden ratio**.

$\square$

```
# plot of F_n / F_(n-1)
points([(n, fibonacci(n)/fibonacci(n-1)) for n in range(2,
    50)])
```

To calculate limits, we can use a version of the limit laws.

Theorem 9.1.11 Sequential Limit Laws. *Let $\{a_n\}$ and $\{b_n\}$ be sequences with $a_n \to a$ and $b_n \to b$. Let c be a constant. Then the following are true:*

1. $a_n + b_n \to a + b$.
2. $ca_n \to ca$.
3. $a_n b_n \to ab$.
4. $\frac{a_n}{b_n} \to \frac{a}{b}$ *assuming* $b \neq 0$.
5. $a_n^p \to a^p$ *if* $a_n, p > 0$.
6. $f(a_n) \to f(a)$ *if* f *is continuous at* a.

Another useful tool for evaluating limits of recursive sequences is the following result: if $\lim_{n\to\infty} a_n = L$, then $\lim_{n\to\infty} a_{n-1} = L$ also.

Example 9.1.12 A Limit from Newton's Method. Find the limit of the sequence

$$x_n = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right) \quad \text{where} \quad x_0 = 1.$$

Solution. First, assume $\lim_{n\to\infty} x_n = x$. Then taking the limit of both sides of the recurrence relation gives

$$x = \frac{1}{2}\left(x + \frac{2}{x}\right).$$

Solving for x, we get $x^2 = 2$, which simplifies to $x = \sqrt{2}$. $\square$

Example 9.1.13 A False Limit. Find the limit of the sequence

$$s_n = 3 - s_{n-1} \quad \text{where} \quad s_0 = 1.$$

Solution. If we let $s = \lim_{n\to\infty} s_n$ and take the limit of both sides of the recurrence, we get $s = 3 - s$ or just $s = \frac{3}{2}$. However, the actual terms of the sequence are given by

$$\{s_n\}_{n=0}^{\infty} = \{1, 2, 1, 2, 1, 2, \ldots\},$$

which is clearly divergent! The problem here is that we assumed a limit existed in the first place. *This is not always valid.* So we need to be careful. □

We can check whether or not a sequence is convergent without actually finding a limit, at least in certain cases.

Theorem 9.1.14 Absolute Value Test. *Suppose that* $\lim_{n\to\infty} |a_n| = 0$*. Then* $a_n \to 0$ *as well.*

Theorem 9.1.15 The Squeeze Theorem for Sequences. *Let* $\{a_n\}, \{b_n\}$ *and* $\{c_n\}$ *be sequences such that*

$$a_n \le b_n \le c_n.$$

If $a_n, c_n \to L$*, then* $b_n \to L$*.*

Example 9.1.16 Applying the Squeeze Theorem and the Absolute Value Test. Let

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

Find $\lim_{n\to\infty} a_n$.

Solution. This sequence is complicated, so we'll try comparing with simpler sequences instead. First, we'll take the absolute value to get rid of the $(-1)^n$ term:

$$|a_n| = \frac{|\sin n|}{\left|\frac{n}{2}\cos n + n\right|} = \frac{|\sin n|}{n\left|1 + \frac{1}{2}\cos n\right|}.$$

Now we'll use the fact that $|a_n| \ge 0, |\sin n| \le 1$ and $\left|1 + \frac{1}{2}\cos n\right| \ge \frac{1}{2}$ to write

$$0 \le |a_n| \le \frac{1}{\frac{n}{2}} = \frac{2}{n}.$$

Since $\frac{2}{n} \to 0$, this forces $|a_n|$, and this a_n, to converge to 0 as well. □

Another important way to check if a sequence converges is the **Monotone Convergence Theorem** Theorem 9.1.19.

Definition 9.1.17 Monotone Sequences. Let $\{a_n\}$ be a sequence. Then $\{a_n\}$ is **increasing** if $a_{n+1} > a_n$ for all n and **decreasing** if $a_{n+1} < a_n$ for all n. In either case, we say that the sequence is monotone. ♢

If we add one more condition to a monotone sequence, we get a convergent sequence.

Definition 9.1.18 Bounded Sequences. Let $\{a_n\}$ be a sequence. We say that $\{a_n\}$ is **bounded** if there exists some real number M such that $|a_n| \le M$ for all n. ♢

Theorem 9.1.19 Monotone Convergence Theorem. *Let* $\{a_n\}$ *be a bounded monotone sequence. Then* $\{a_n\}$ *converges.*

Example 9.1.20 Applying the MCT. Let $\{a_n\}$ denote the sequence

$$\{\sqrt{2}, \sqrt{2\sqrt{2}}, \sqrt{2\sqrt{2\sqrt{2}}}, \ldots\}$$

Determine if the sequence converges and if so find its limit.

Solution. First, note that

$$a_n = \sqrt{2a_{n-1}} \quad \text{and} \quad a_1 = \sqrt{2}.$$

To show this converges, we'll use the MCT. To do so, we must show that the sequence is bounded and increasing. To show it's bounded, we'll guess that

$a_n < 2$ for some n. Then

$$a_{n+1} = \sqrt{2a_n} < \sqrt{2}\sqrt{2} = 2,$$

implying the claim. Now,

$$a_{n+1} = \sqrt{2}\sqrt{a_n} > \sqrt{a_n^2} = a_n,$$

showing the sequence is increasing. Hence it's convergent by the MCT. The limit is equal to 2. □

9.2 Series

Consider the number $\pi = 3.14159\ldots$. This number is irrational and so cannot be represented as a rational number $\frac{a}{b}$. This leads to the question of what we mean by π? Or in particular, how can we actually make sense of π, or represent it?

We can consider rewriting π as follows:

$$\pi = 3.14159\ldots = 3 + \frac{1}{10} + \frac{4}{100} + \frac{1}{1000} + \frac{5}{10000} + \frac{9}{100000} + \cdots.$$

So we can identify π with the sequence $\{a_n\}_{n=0}^{\infty} = \{3, \frac{1}{10}, \frac{4}{100}, \ldots\}$ and the **series**

$$a_0 + a_1 + a_2 + \cdots.$$

Definition 9.2.1 Infinite Series. An infinite series is a sum of the form

$$\sum_{k=0}^{\infty} a_k = a_0 + a_1 + \cdots$$

where $\{a_k\}_{k=0}^{\infty}$ is a sequence. ◊

Infinite series are useful for representing (and computing) irrational numbers (which includes almost all numbers).

Example 9.2.2 Guessing Sums. Guess the sums of the following series:

1. $\displaystyle\sum_{k=0}^{\infty} 2^k$
2. $\displaystyle\sum_{k=1}^{\infty} 2^{-k}$
3. $1 - 1 + 1 - 1 + 1 - \cdots$

Solution. We have the following:

1. $\displaystyle\sum_{k=0}^{\infty} 2^k = \infty$
2. $\displaystyle\sum_{k=1}^{\infty} 2^{-k} = 1$
3. For this last sum we have an issue: there's no sensible way to define this sum. We can say that $1 - 1 + \cdots = 0$ or 1 by grouping terms differently.

□

We can determine what the value of a series should be by using limits.

Definition 9.2.3 Partial Sums and Convergence. Given the series $\sum_{k=0}^{\infty} a_k$, we denote its n^{th} **partial sum** by

$$S_n = \sum_{k=0}^{n} a_k.$$

If the sequence $\{S_n\}$ is convergent and $S_n \to S$, then we say that the original series **converges** and $\sum_{k=0}^{\infty} a_k = S$. If the sequence of partial sums diverges, we say the original series **diverges**. $\Diamond$

Using Definition 9.2.3, we can say that the sum $1 - 1 + 1 - 1 + \cdots$ diverges, since its sequence of partial sums is $\{1, 0, 1, 0, 1, \ldots\}$. The same is true for the first series in Example 9.2.2, but the second series converges.

Example 9.2.4 Determining Convergence of a Series. Does the series $\sum_{n=1}^{\infty} 3^{-n}$ converge?

Solution. We'll look at the sequence of partial sums. We have

$$S_1 = \frac{1}{3}$$
$$S_2 = \frac{4}{9}$$
$$S_3 = \frac{13}{27}$$

and so on. It looks like the sequence of partial sums approaches $\frac{1}{2}$, so we guess that the series equals the same. $\square$

The series in Example 9.2.4, as well as the first two series in Example 9.2.2, are examples of an important series known as a **geometric series**.

Definition 9.2.5 Geometric Series. A series $\sum_{k=0}^{\infty} a_k$ is a geometric series if $a_k = ar^k$ for some constants a and r. Equivalently, the terms of the series form a geometric sequence (see Example 9.1.7). $\Diamond$

Geometric series are useful because it's straightforward to find their values. To see how, let $\sum_{k=0}^{\infty} ar^k$ be a geometric series and let $\{S_n\}$ denote the corresponding sequence of partial sums. Then

$$S_n = a + ar + ar^2 + \cdots + ar^n$$

which gives

$$S_n - rS_n = a - ar^{n+1}.$$

We can solve this for S_n to get

$$S_n = \frac{a - ar^{n+1}}{1 - r}.$$

At this point, we can find the limit of the partial sums using Example 9.1.7. Therefore $\sum_{k=0}^{\infty} ar^k$ converges to $\frac{a}{1-r}$ if $|r| < 1$ and diverges otherwise.

As a quick example of this result, we can find the value of $\sum_{n=1}^{\infty} 3^{-n}$ since this series is geometric. To do so, we must determine a and r for this sum. Since

$$\sum_{n=1}^{\infty} 3^{-n} = \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots,$$

we have $a = \frac{1}{3}$ and $r = \frac{1}{3}$ also. Hence the series sums to $\frac{1/3}{2/3} = \frac{1}{2}$.

Example 9.2.6 Computing a Geometric Series. Determine the value of $\sum_{k=0}^{\infty}\left(-\frac{1}{4}\right)^k 5^{6-k}$ if it exists.

Solution. Since this series contains terms being raised to the k^{th} power, we suspect it may be geometric. If we write out the first several terms, we get

$$\sum_{k=0}^{\infty}\left(-\frac{1}{4}\right)^k 5^{6-k} = 5^6 - \frac{5^5}{4} + \frac{5^4}{4^2} - \cdots,$$

so at each step we're dividing by $4\cdot 5 = -20$. This series is therefore a geometric series with $a = 5^6$ and $r = -\frac{1}{20}$. Since $|r| < 1$, this series converges. The value of this series is

$$\sum_{k=0}^{\infty}\left(-\frac{1}{4}\right)^k 5^{6-k} = \frac{a}{1-r} = \frac{5^6}{21/20}.$$

□

We can also find a and r without writing out the first few terms of the series.

Example 9.2.7 Finding a and r. Determine the value of $\sum_{k=2}^{\infty} 3^{4-k}7^{3k}$.

Solution. We can rewrite the series as

$$\sum_{k=2}^{\infty} 3^{4-k}7^{3k} = \sum_{k=2}^{\infty} 3^4\left(\frac{7^3}{3}\right)^k.$$

This is a geometric series with

$$a = 3^4\frac{7^6}{3^2} \quad \text{and} \quad r = \frac{7^3}{3}.$$

Since $|r| > 1$, the series diverges. □

Example 9.2.8 $.9\ldots$ repeating. Prove that $.9\ldots = 1$ using geometric series.

Solution. First, we need to write $.9\ldots$ as a geometric series. We can do so as follows:

$$.9\ldots = \frac{9}{10} + \frac{9}{100} + \cdots,$$

and so we see that

$$.9\ldots = \sum_{k=1}^{\infty}\frac{9}{10^k}.$$

This is a geometric series with $a = \frac{9}{10}$ and $r = \frac{1}{10}$ (and so is convergent!), and so

$$\sum_{k=1}^{\infty}\frac{9}{10^k} = \frac{9/10}{1-\frac{1}{10}} = 1.$$

□

Example 9.2.9 Writing a Decimal as a Fraction. Rewrite the decimal $.123451234512345\ldots$ as a fraction $\frac{m}{n}$.

Solution. First, it's a mathematical fact that any repeating decimal can be written as a rational number so we know that we can actually write $0.\overline{12345}$ as a fraction. We'll do so by rewriting the decimal as a geometric series:

$$0.\overline{12345} = \frac{12345}{100000} + \frac{12345}{10000000000} + \cdots,$$

which is a geometric series with $a = \frac{12345}{100000}$ and $r = \frac{1}{100000}$. This series is also convergent, and has sum

$$\frac{a}{1-r} = \frac{12345/100000}{99999/100000} = \frac{12345}{99999}.$$

□

Another type of series that can be calculated (relatively) easily is the **telescoping series**. We'll demonstrate by way of example.

Example 9.2.10 Telescoping Logarithms. Find $\sum_{k=1}^{\infty} \ln \frac{k+1}{k}$.

Solution. If we write out the first few terms, we get

$$\sum_{k=1}^{\infty} \ln \frac{k+1}{k} = (\ln 2 - \ln 1) + (\ln 3 - \ln 2) + \cdots$$

so it looks like many of these terms cancel each other out. To be precise about this, we'll find the partial sums of this series and then consider their limit:

$$S_n = \sum_{k=1}^{n} \ln \frac{k+1}{k} = \ln(n+1),$$

which goes to ∞ as $n \to \infty$. So the series diverges. □

Not every series is obviously a telescoping series.

Example 9.2.11 Rewriting a Telescoping Series. Find $\sum_{n=3}^{\infty} \frac{3}{n(n-1)}$.

Solution. It's not obvious at all that the series is telescoping, even if we write out a few terms. However, if we try partial fractions on $\frac{3}{n(n-1)}$ we obtain (see SageMath cell below)

$$\frac{3}{n(n-1)} = \frac{3}{n-1} - \frac{3}{n}.$$

So

$$\sum_{n=3}^{\infty} \frac{3}{n(n-1)} = \sum_{n=3}^{\infty} \left[\frac{3}{n-1} - \frac{3}{n}\right].$$

The k^{th} partial sum is

$$S_k = \sum_{n=3}^{k} \left[\frac{3}{n-1} - \frac{3}{n}\right] = \frac{3}{2} - \frac{3}{k},$$

and so $S_k \to \frac{3}{2}$. Hence $\sum_{n=3}^{\infty} \frac{3}{n(n-1)} = \frac{3}{2}$. □

```
# Finds partial fraction decomposition for previous example
import sympy as sp
sp.init_printing()
n = sp.Symbol('n')

expr = 3 / ( n * ( n - 1 ) )
expr.apart()
```

The next series is important despite diverging.

Example 9.2.12 The Harmonic Series. Show that the **harmonic series**

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \cdots$$

is divergent.

Solution. The idea here (which will return later) is to compare this series with a simpler one that we know diverges. We'll do so by looking at a specific set of partial sums:

$$\begin{aligned} S_1 &= 1 \\ S_2 &= 1 + \frac{1}{2} \\ S_4 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} > 1 + \frac{1}{2} + \frac{1}{2} \\ S_8 &> 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \end{aligned}$$

and in general

$$S_{2^n} > 1 + \frac{n}{2}.$$

So it follows that

$$\lim_{n\to\infty} S_n \geq \lim_{n\to\infty} \left(1 + \frac{n}{2}\right) = \infty.$$

Hence the harmonic series is divergent. □

A useful test for divergence of a series involves the long-term behavior of the terms of the series.

Theorem 9.2.13 Divergence Test. *Consider the series* $\sum a_k$*. If* $\lim_{k\to\infty} a_k \neq 0$*, then* $\sum a_k$ *diverges.*

Proof. We'll prove the **contrapositive** of this statement. That is, we'll show that if the series converges then the terms go to 0. So suppose $\sum a_k$ converges and let S_n denote the sequence of partial sums. Then $a_{n+1} = S_{n+1} - S_n$ which must go to 0 since the partial sums converge. ■

Note that Theorem 9.2.13 cannot be used to prove *convergence*, only divergence. For example, the terms of the harmonic series go to 0 but the series itself diverges.

Example 9.2.14 Using the Divergence Test. Determine if $\sum_{n=1}^{\infty} \sin n$ diverges.

Solution. Since $\lim_{n\to\infty} \sin n \neq 0$ (in fact, it doesn't exist at all), the series must diverge. □

Series, or rather the summation symbol Σ, obey many of the same laws as integrals: they split over sums and we may pull constants out.

Example 9.2.15 Splitting a Sum. Find the value of $\sum_{k=4}^{\infty} \left(3 \cdot 2^{-k} - \pi e^{-k}\right)$. □

9.3 The Integral and Comparison Tests

Convergence Tests. The Divergence Test proven in Theorem 9.2.13 is our first example of a **convergence test**: a test that determines if a given series converges or diverges. In this section we'll introduce two more such tests. It's important to remember that convergence tests *usually cannot be used to evaluate a series*. Their primary importance is to check if a given series converges.

The Integral Test

The main idea behind the **integral test** is to relate the value of a series to the value of a certain (improper) integral. This is useful since integrals are often easier to compute than series.

Theorem 9.3.1 Integral Test. *Suppose that $f(x)$ is a positive, decreasing function on $[1,\infty)$ and let $a_n = f(n)$. Then*

$$\int_1^\infty f(x)\,dx \quad \textit{and} \quad \sum_{n=1}^\infty a_n$$

must both converge or both diverge.

Remark 9.3.2 Remember that the Integral Test usually *cannot* determine the value of a series. It can only be used to determine convergence.

Example 9.3.3 Determining Convergence Using the Integral Test. Determine if the series $\sum_{k=0}^\infty \frac{1}{1+k^2}$ converges or diverges.

Solution. We can use the Integral Test here since $\frac{1}{1+k^2}$ is positive and decreasing. If we define $f(x) = \frac{1}{1+x^2}$, then $\frac{1}{1+k^2} = f(k)$. Now we'll compute $\int_0^\infty f(x)\,dx$:

$$\begin{aligned}\int_0^\infty \frac{1}{1+x^2}\,dx &= \lim_{b\to\infty}\int_0^b \frac{1}{1+x^2}\,dx \\ &= \lim_{b\to\infty} \tan^{-1} b \\ &= \frac{\pi}{2}.\end{aligned}$$

Since the integral converges, so does the series. In fact, the value of the series is

$$\frac{\pi\coth\pi + 1}{2}.$$

□

Example 9.3.4 The Alternating Harmonic Series. Explain why Theorem 9.3.1 cannot be applied to the **alternating harmonic series**

$$\sum_{k=1}^\infty \frac{(-1)^{k+1}}{k}.$$

Solution. Since $\frac{(-1)^{k+1}}{k}$ is neither decreasing nor positive, the Integral Test doesn't apply here. □

An important corollary to Theorem 9.3.1 is that the integral p-test from Section 7.4 applies to series as well.

Theorem 9.3.5 Series p-Test. *The series $\sum_{n=1}^\infty \frac{1}{n^p}$ converges if and only if $p > 1$.*

Comparison Tests

Theorem 9.3.6 Comparison Test. *Let $\sum a_k$ and $\sum b_k$ be series with* positive *terms. Then*

1. *If $a_k \leq b_k$ and $\sum b_k$ converges, then so does $\sum a_k$.*

2. *If $a_k \geq b_k$ and $\sum b_k$ diverges, then so does $\sum a_k$.*

Example 9.3.7 Using the Comparison Test. Show that $\sum_{k=1}^{\infty} \frac{k+1}{k^2}$ diverges. □

Sometimes using the Comparison Test requires a little ingenuity.

Example 9.3.8 A Little Ingenuity. Show that $\sum_{k=1}^{\infty} \frac{k^2+k-1}{k^4+4k^2-3}$ converges. □

A test that is sometimes more straightforward is the **Limit Comparison Test**.

Theorem 9.3.9 Limit Comparison Test. *Suppose that $\sum a_k$ and $\sum b_k$ are both series with positive terms, and suppose*

$$L = \lim_{k \to \infty} \frac{a_k}{b_k}$$

exists. Then

1. *if $0 < L < \infty$, then either both series converge or both series diverge.*

2. *if $L = 0$ and $\sum b_k$ converges, then so does $\sum a_k$.*

3. *if $L = \infty$ and $\sum b_k$ diverges, then so does $\sum a_k$.*

The quantity L in Theorem 9.3.9 can be thought of as the relative size of a_k as compared to b_k.

Example 9.3.10 A Little Less Ingenuity. Show that $\sum_{k=1}^{\infty} \frac{k^2+k-1}{k^4+4k^2-3}$ converges.

Solution. We saw previously that

$$\frac{k^2+k-1}{k^4+4k^2-3} \approx \frac{1}{k^2},$$

which suggests comparing the original series with the p-series $\sum \frac{1}{k^2}$. If we let $a_k = \frac{k^2+k-1}{k^4+4k^2-3}$ and $b_k = \frac{1}{k^2}$, then we see that

$$\frac{a_k}{b_k} \to 1.$$

By Theorem 9.3.9 and Theorem 9.3.5, the original series converges. □

The Limit Comparison Test works very well with series containing terms given by a ratio of powers, in conjunction with the p-series Test.

Example 9.3.11 Radical Powers of k. Does

$$\sum_{k=55}^{\infty} \frac{k^3 - 3\sqrt{k} + 1}{\sqrt[4]{k^{10}} + k^2}$$

converge or diverge?

Solution. The series diverges by comparison with $\frac{1}{k^{1/3}}$. □

9.4 Other Convergence Tests

Alternating Series

An **alternating series** is any series whose terms switch sign. Written in summation notation, they take the form $\sum(-1)^k a_k$ where a_k is a positive sequence. Alternating series have a very useful test for convergence.

Theorem 9.4.1 Alternating Series Test. *Consider the alternating series $S = \sum(-1)^k a_k$, where a_k is positive* and decreasing. *If $a_k \to 0$, then the series converges. Furthermore, for such a series we have the* ***remainder*** *estimate*

$$|R_n| = |S - S_n| \leq |a_{n+1}|.$$

Note that Theorem 9.4.1 is *not* the same as Theorem 9.2.13.

Example 9.4.2 An Alternating Series with Roots. Does

$$\sum_{n=6}^{\infty}(-1)^n \frac{n^{\frac{n+1}{n}} - n}{n}$$

converge or diverge?

Solution. Let $a_n = \frac{n^{\frac{n+1}{n}}-n}{n} = n^{1/n} - 1$. Then a_n is decreasing, and $a_n \to 0$, so the series converges. □

Example 9.4.3 Alternating Harmonic Series. Show that the Alternating Harmonic Series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$ converges, and determine a value of n for which S_n is within .001 of the actual value of $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$.

Solution. Since the Alternating Harmonic Series is an alternating series with $a_k = \frac{1}{k}$, and because these terms decrease to 0, the sum must converge. However, we do not yet know *what* it converges to yet. Now let S_n denote the n^{th} partial sum. Then we know the error between S and S_n is at most

$$|a_{n+1}| = \frac{1}{n+1}.$$

To make this less than .001, we must have $n \geq 999$. □

The Alternating Harmonic Series is also a useful example to illustrate the following definitions.

Definition 9.4.4 Absolute and Conditional Convergence. A series $\sum a_k$ is **absolutely convergent** if $\sum |a_k|$ converges. A series $\sum a_k$ is **conditionally convergent** if it converges but $\sum |a_k|$ diverges. ◊

The Alternating Harmonic Series is an example of a conditionally convergent series. There are two important consequences of Definition 9.4.4:

1. Absolutely convergent series are also convergent series.

2. For conditionally convergent series, *order matters*.

Example 9.4.5 Convergence of a Series Involving Sine. Determine if

$$\sum_{n=4}^{\infty} \frac{\sin(3n - \sqrt{n})}{3^n}$$

converges or diverges.

Solution. If we take the absolute value of each term, then we get

$$\frac{|\sin(3n - \sqrt{n})|}{3^n} \leq \frac{1}{3^n}.$$

Since $\sum_{n=4}^{\infty} 3^{-n}$ is a geometric series with $|r| < 1$, then

$$\sum_{n=4}^{\infty} \left| \frac{|\sin(3n - \sqrt{n})|}{3^n} \right|$$

must converge by Theorem 9.3.6.

Hence the original series is absolutely convergent, and so also convergent.

□

Ratio Test

Geometric series are among the easiest to sum and determine convergence for. So it's useful to try to compare an arbitrary series with a geometric series. The main idea is to look at the long-term behavior of *ratios* of consecutive terms.

Theorem 9.4.6 Ratio Test. *Let $\sum a_k$ be an infinite series and let*

$$r = \lim_{k\to\infty} \left| \frac{a_{k+1}}{a_k} \right|.$$

Then

1. *If $r < 1$ the series converges absolutely.*
2. *If $r > 1$ the series diverges.*
3. *If $r = 1$ the test fails.*

Example 9.4.7 Using the Ratio Test. Does

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1} 3^{2k+1}}{k^2 5^k}$$

converge or diverge?

Solution. Since

$$r = \frac{9}{5} > 1,$$

the series diverges by Theorem 9.4.6. □

The ratio test works well with series whose terms involve factorials or powers involving k.

Example 9.4.8 Factorials over Powers. Show that $\sum_{k=0}^{\infty} \frac{k!}{k^k}$ converges.

Solution. Since $a_k = \frac{k!}{k^k}$, we have

$$\lim_{k\to\infty} \left| \frac{a_{k+1}}{a_k} \right| = \lim_{k\to\infty} \left(\frac{k}{k+1} \right)^k.$$

We can find this limit using L'Hospital's Rule (see Section 4.6) since this limit is the indeterminate form $[1^\infty]$. So set $y = \lim_{k\to\infty} \left(\frac{k}{k+1}\right)^k$. Then

$$\ln y = \lim_{k\to\infty} k \ln\left(\frac{k}{k+1} \right)$$

$$
\begin{aligned}
&= \lim_{k\to\infty} \frac{\ln k - \ln(k+1)}{1/k} \\
&= \lim_{k\to\infty} \frac{1/k - 1/(k+1)}{-1/k^2} \\
&= \lim_{k\to\infty} \frac{-k^2}{k(k+1)} \\
&= -1.
\end{aligned}
$$

Therefore $y = \lim_{k\to\infty} \left(\frac{k}{k+1}\right)^k = e^{-1}$, which means that the series converges by the ratio test. □

Root Test

The root test is similar to the ratio test in that it compares a given series with an appropriate geometric series to determine if the original converges.

Theorem 9.4.9 Root Test. *Let $\sum a_k$ be an infinite series and let*

$$r = \lim_{k\to\infty} \sqrt[k]{|a_k|}.$$

Then

1. *If $r < 1$ the series converges absolutely.*
2. *If $r > 1$ the series diverges.*
3. *If $r = 1$ the test fails.*

Example 9.4.10 A $k^{\mathbf{th}}$ Power. Show that $\sum_{k=3}^{\infty} \frac{2^k}{k^{10}}$ diverges. □

Example 9.4.11 A Series with Rational Terms. Does $\sum_{k=5}^{\infty} (-1)^{2k-3} \left(\frac{3k^2-5k}{k^3-1}\right)^{2k}$ converge or diverge? □

9.5 Power Series

A **power series** is a series of the form

$$\sum_{k=0}^{\infty} a_k x^k$$

where x is a variable. Note that for such a series, only nonnegative, integer powers of x are permitted. The terms a_k are called **coefficients**, and we'll see later that they determine all properties of the series.

Example 9.5.1 Examples of Power Series. Determine which of the following are power series:

1. $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!}$
2. $x - \pi x^4 + x^6 + \dfrac{1}{10} x^{100} + \cdots$
3. $\dfrac{1}{x^2} + \dfrac{1}{x} + 1 + x + x^2 + \cdots$

□

Power series can also be *centered* at other numbers. A power series **centered about a** is a series of the form

$$\sum_{k=0}^{\infty} a_k (x-a)^k.$$

An important concern about power series is for which values of x the series will converge. These questions are usually answered using the root or ratio tests.

Example 9.5.2 Convergence of a Power Series. For what values of x does the series

$$\sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1}$$

converge?

Solution. We'll try the ratio test to check convergence of this series. Doing so, we get

$$r = \lim_{k\to\infty} |x| \frac{k+1}{k+2} = |x|.$$

So the series converges if $|x| < 1$ and diverges if $|x| > 1$.

When $r = |x| = 1$, or $x = \pm 1$, the test fails. So we need to use other methods to determine the convergence or the series at these points. At $x = -1$, the series becomes

$$\sum_{k=0}^{\infty} \frac{(-1)^{2k+1}}{k+1} = -\sum_{k=0}^{\infty} \frac{1}{k+1},$$

which diverges by the comparison test. At $x = 1$, the series reduces to the alternating harmonic series which converges by Example 9.4.3.

Therefore this series converges for all x in the interval $(-1, 1]$ and diverges otherwise. □

In Example 9.5.2, the values of x for which the series converged was an interval. It turns out that this will always be the case, and the resulting interval is known as the **interval of convergence** of the series. The "radius" of this interval is called the **radius of convergence**. In general, we have the following.

Theorem 9.5.3 Convergence of Power Series. *Given a series* $\sum_{k=0}^{\infty} a_k x^k$*, there exists* $R \geq 0$ *such that the series converges on the interval* $(a - R, a + R)$*. The largest such* R *is the radius of convergence.*

For most series we'll consider (i.e., those of the form $\sum a_k (x-a)^k$), we can find R using the following formula:

$$R = \lim_{k\to\infty} \left| \frac{a_k}{a_{k+1}} \right|.$$

Example 9.5.4 Interval and Radius of Convergence. Find the interval and radius of convergence of the series $\sum_{k=0}^{\infty} \frac{(-1)^{2k}(x-\frac{\pi}{2})^{2k}}{(2k)!}$.

Solution. We'll find the radius of convergence first, which is given by

$$\begin{aligned} R &= \lim_{k\to\infty} \frac{(2k+2)!}{(2k)!} \\ &= \lim_{k\to\infty} (2k+2)(2k+1) \\ &= \infty. \end{aligned}$$

So the radius of convergence is infinite, implying that the interval of convergence is $(-\infty, \infty)$. □

We can also use the root test instead of the ratio test.

Example 9.5.5 Interval and Radius of Convergence from Root Test. Determine the interval and radius of convergence of

$$\sum_{n=1}^{\infty} \frac{(-2)^n}{\sqrt{n}}(4x+3)^n.$$

Solution. If we apply the root test to this series, we get

$$r = \lim_{n\to\infty} \sqrt[n]{\frac{2^n|4x+3|^n}{\sqrt{n}}} = \lim_{n\to\infty} \frac{2}{n^{1/2n}}|4x+3| = 2|4x+3|.$$

We need this to be less than 1, which gives

$$2|4x+3| < 1 \Rightarrow -\frac{7}{8} < x < -\frac{5}{8},$$

and so the series converges for all x in $(-\frac{7}{8}, -\frac{5}{8})$. So the radius of convergence is $\frac{1}{8}$.

Now we need to check the endpoints. At $x = -\frac{7}{8}$, the series becomes $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$, which diverges by the p-series test. At $x = -\frac{5}{8}$, the series becomes $\sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}}$ which converges by the alternating series test.

Therefore the interval of convergence is $(-\frac{7}{8}, -\frac{5}{8}]$. □

9.6 Representing Functions as Power Series

A power series can be viewed as a function with domain given by the interval of convergence. One of our goals is to use power series to represent functions that can't be written in terms of "elementary functions". Our first example comes from the geometric series.

Example 9.6.1 Power Series from Geometric Series. Let

$$f(x) = \sum_{k=0}^{\infty} x^k.$$

Then the domain of $f(x)$ is $(-1, 1)$ and for x in this interval we have

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

□

Note that the equation

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

is only valid where the series on the left converges. If we try to plug in $x = -1$ into this equation and treat it as valid, we get

$$\frac{1}{2} = 1 - 1 + 1 - 1 + \cdots,$$

which is nonsense.

Now that we have a power series representation of $\frac{1}{1-x}$, we can use it to find other power series representations.

Example 9.6.2 Finding a Power Series Representation. Find a power series representation for $\dfrac{x}{1+4x}$ and state the interval over which it is valid.

Solution. We'll try to relate this to the series in Example 9.6.1:

$$\begin{aligned}\frac{x}{1+4x} &= x\frac{1}{1+4x}\\ &= x\frac{1}{1-(-4x)}\\ &= x\sum_{k=0}^{\infty}(-4x)^k \quad \text{valid for } |4x|<1\\ &= \sum_{k=0}^{\infty}(-1)^k 4^k x^{k+1}.\end{aligned}$$

This representation is valid for $|x| < \frac{1}{4}$, or for x in the interval $(-\frac{1}{4}, \frac{1}{4})$. □

One of the most useful properties of power series is that the fundamental calculus operations, differentiation and integration, are valid for power series within their intervals of convergence.

Theorem 9.6.3 Differentiation and Integration of Power Series. *Suppose the power series $\sum_{k=0}^{\infty} a_k(x-a)^k$ has radius of convergence $R > 0$, and let $f(x)$ denote the series on the interval $(a-R, a+R)$. Then $f(x)$ is differentiable on $(a-R, a+R)$ and*

$$\begin{aligned}f'(x) = \frac{d}{dx}&\left(a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + \cdots\right)\\ &a_1 + 2a_2(x-a) + 3a_3(x-a)^2 + \cdots\end{aligned}$$

or in other words

$$f'(x) = \frac{d}{dx}\sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} k a_k x^{k-1}.$$

Similarly,

$$\int f(x)\,dx = \int \sum_{k=0}^{\infty} a_k x^k\,dx = \sum_{k=0}^{\infty} a_k \frac{x^{k+1}}{k+1} + C.$$

Both of these series have radius of convergence R.

Theorem 9.6.3 shows that integrating and differentiating power series is as easy as integrating or differentiating powers of x. However, we do need to be careful at the endpoints.

Example 9.6.4 Power Series for the Logarithm. Find a power series representation of $\ln(1+x)$ centered at 0 and its interval of convergence.

Solution. Since $\dfrac{d}{dx}\ln(1+x) = \frac{1}{1+x}$, we can integrate the series for $\frac{1}{1+x}$ to get the series for the logarithm. Doing so, we get

$$\ln(1+x) + C = \int \sum_{k=0}^{\infty}(-1)^k x^k\,dx$$

$$= \sum_{k=0} (-1)^k \frac{x^{k+1}}{k+1}.$$

The series on the right has radius of convergence $R = 1$ and interval of convergence $(-1, 1]$ by Example 9.5.2. To find C, we can substitute $x = 0$ into the equation (which is valid!) to get

$$\ln 1 + C = 0 - \frac{0^2}{2} + \cdots = 0.$$

So $C = 0$, and

$$\ln(1 + x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1}.$$

Plugging in $x = 1$, we get the series

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \cdots.$$

□

9.7 Taylor and Maclaurin Series

In Section 9.6 we found how to obtain power series using the geometric series. What we'll do in this section is determine how to find more general power series. To start, suppose we can represent $f(x)$ as a power series centered at some number a:

$$f(x) = \sum_{k=0}^{\infty} c_k (x-a)^k = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots$$

Then this representation is valid on some interval $(a - R, a + R)$. Our goal is to find what the coefficients c_k are.

We'll start with the first coefficient c_0. We can find it very easily by setting $x = a$ to get

$$f(a) = c_0.$$

So the first coefficient *must* be equal to $f(a)$. Now we'll try to find c_1, which we can do by differentiating:

$$f'(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots,$$

which is still valid for $(a - R, a + R)$. Now set $x = a$ again to obtain

$$f'(a) = c_1.$$

Now differentiate again to get

$$f''(x) = 2a_2 + 3 \cdot 2(x-a) + \cdots$$

which gives

$$a_2 = \frac{f''(a)}{2}.$$

Similarly,

$$a_3 = \frac{f'''(a)}{3!}.$$

We can continue this process in general to get the following theorem.

Theorem 9.7.1 Taylor's Formula. *If $f(x)$ has a power series representation/expansion at a with positive radius of convergence R, then*

$$f(x) = \sum_{k=0}^{\infty} c_k (x-a)^k$$

where

$$c_k = \frac{f^{(k)}(a)}{k!}.$$

This series is called the ***Taylor series*** *of f about a.*

A Taylor series about 0 is also known as a **Maclaurin series**.

Example 9.7.2 Maclaurin Series for the Exponential. Find the Maclaurin series for e^x.

Solution. If we let $f(x) = e^x$, then the Maclaurin series is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$$

where

$$f^{(k)}(0) = e^0 = 1.$$

So the Maclaurin series is

$$\sum_{k=0}^{\infty} \frac{1}{k!} x^k = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \cdots.$$

□

Example 9.7.2 gives us the Taylor series about 0 for e^x, but at the moment we don't know if it actually equals e^x. However, if we differentiate the series we get the same series back. In other words, the Taylor series of e^x *is its own derivative*, which is a promising sign.

Example 9.7.3 Taylor Series from the Exponential. Assuming that e^x equals its Taylor series, find the Maclaurin series for e^{-x^2}. Also, find the 70^{th} derivative of e^{-x^2} at $x = 0$.

Solution. We could use Theorem 9.7.1 to find the Maclaurin series for e^{-x^2}, but it's far, far easier to use the series for e^x:

$$e^{-x^2} = \sum_{k=0}^{\infty} \frac{(-x^2)^k}{k!} = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{k!}.$$

It turns out that this lets us find derivatives at 0 incredibly quickly. If we let $f(x) = e^{-x^2}$, then it follows that $\frac{f^{(k)}(0)}{k!}$ is the coefficient of x^k in the power series for $f(x)$. Therefore

$$\frac{f^{(70)}(0)}{70!} = \frac{(-1)^{35}}{35!} \Rightarrow f^{(70)}(0) = -\frac{70!}{35!}.$$

□

Showing that a function equals its power series requires the **Lagrange remainder formula** for Taylor series, which states the following: if $f(x)$ is differentiable $N+1$ times on an interval containing a, then for each x in I

there exists z in I such that

$$R_N(x) = f(x) - \sum_{k=0}^{N} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

is equal to

$$\frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1}.$$

A useful fact when using this is that

$$\frac{x^N}{N!} \to 0$$

as $N \to \infty$ for all x.

Example 9.7.4 Maclaurin Series for Cosine. Find the Maclaurin series for $\cos x$ and show that $\cos x$ equals its Maclaurin series for all x.

Solution. Using Theorem 9.7.1, the Taylor series for $f(x) = \cos x$ at $a = 0$ is given by

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k.$$

So to find this series we need to find the derivatives of cosine at 0:

$$\begin{aligned} f(0) &= 1 \\ f'(0) &= 0 \\ f''(0) &= -1 \\ f^{(3)}(0) &= 0 \end{aligned}$$

and so on. So the Maclaurin series for $\cos x$ is given by

$$1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \cdots.$$

To express this using sigma notation, note that this series has only even powers of x. Hence the sum should look like $\sum_{k=0}^{\infty} c_k x^{2k}$. Looking at the pattern of the coefficients here, we can rewrite this as

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}.$$

Using the ratio test, we can also show that this series converges for all x. To show that this series actually equals $\cos x$, we need to consider the remainder term

$$R_N(x) = \frac{f^{(N+1)}(z)}{(N+1)!} x^{N+1}.$$

We don't know what exactly z is, except that it's in some interval around 0. However, we do know that $\|f^{(N+1)}(z)\| \leq 1$, which means that the remainder satisfies

$$\|R_N(x)\| \leq \frac{|x|^{N+1}}{(N+1)!} \to 0.$$

Therefore $\cos x$ equals its Maclaurin series. □

Now that we know how to express $\cos x$ as a power series, we can do the same for $\sin x$.

Example 9.7.5 Maclaurin Series for Sine. Find the Maclaurin series for $\sin x$. □

Knowing a power series representation for a function makes calculus with that function extremely straightforward, at least if you're willing to use series.

Example 9.7.6 Computing a Definite Integral. Use power series to find $\displaystyle\int_0^1 e^{-x^2}\,dx$, and find an approximation within .01 of the exact value.

Solution. Using the power series for e^{-x^2} from Example 9.7.3, we have

$$\int_0^1 e^{-x^2}\,dx = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)k!}.$$

This is an alternating series, so by the alternating series test the partial sum $\displaystyle\sum_{k=0}^{N} \frac{(-1)^k}{(2k+1)k!}$ is always within

$$\frac{1}{(2(N+1)+1)(N+1)!}$$

of the exact value. So if we want to get enough terms of the series to be within .01 of the exact value, we need to pick N so that

$$\frac{1}{(2(N+1)+1)(N+1)!} < .01,$$

which occurs at $N = 3$. So

$$\int_0^1 e^{-x^2}\,dx \approx 1 - \frac{1}{3} + \frac{1}{10} - \frac{1}{42} = 0.742....$$

This is within about .004 of the exact value. □

One extremely useful application of power series in mathematics and its applications is in deriving **Euler's Formula**.

Theorem 9.7.7 Euler's Formula. *Let $i = \sqrt{-1}$. Then $e^{ix} = \cos x + i \sin x$.*

Proof. We can try to use power series to make sense of e^{ix}. Our idea is that since the power series for e^x is valid for all x, it should be valid for all ix as well, giving

$$\begin{aligned} e^{ix} 1 + (ix) + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \cdots \\ = \cos x + i \sin x. \end{aligned}$$

In particular,

$$e^{i\pi} = -1.$$

■

Part III

Multivariable Calculus

Chapter 10

Parametric Equations and Polar Coordinates

Up to this chapter we've represented curves in the xy-plane using Cartesian (i.e. xy) coordinates. What we'll do now is introduce a new method of describing curves in the plane by way of parametric equations.

10.1 Parametric Equations

If a particle is moving in two dimensions (i.e., the xy-plane), then it makes sense to write the x- and y-coordinates of its **trajectory** in terms of time t:

$$\begin{aligned} x &= x(t) \\ y &= y(t). \end{aligned}$$

We can use **parametric equations** to describe the resulting **parametric curves**. For example, the parametric equations

$$\begin{aligned} x &= 2t^2 - 1 \\ y &= t + 1 \\ -2 \leq t &\leq 2 \end{aligned}$$

produce the following parametric curve in the xy-plane.

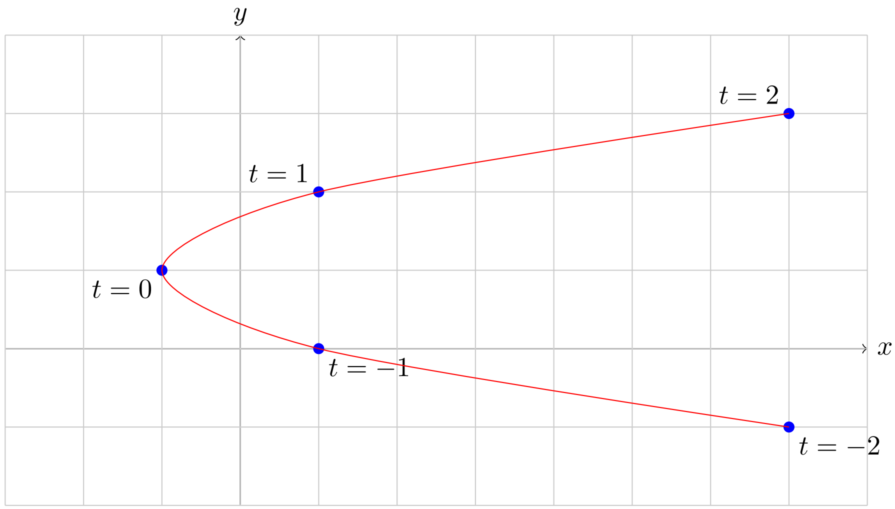


Figure 10.1.1 The graph of $x = 2t^2 - 1, y = t + 1$ for $-2 \leq t \leq 2$.

Note that the parametric curve above has a *starting point* and an *ending point*, determined by the interval of values for t. When plotting parametric curves, we'll often specify a corresponding interval of values for t. This curve can also be plotted relatively painlessly with SymPy, as below:

```
import sympy as sp
from sympy.plotting import plot_parametric

t = sp.symbols('t')
plot_parametric(2 * t**2 - 1, t + 1, (t, -2, 2))
```

A set of parametric equations (along with a corresponding interval of values for t) can be viewed as defining a single function $f(t)$ that assigns points in the xy-plane to real numbers t. We'll introduce some notation that will come in handy later: let $\mathbb{R}$ denote the set of real numbers (also known as **scalars**), and let $\mathbb{R}^2$ denote the xy-plane. Then a set of parametric equations describes a function from $\mathbb{R}$ (specifically, an interval in $\mathbb{R}$) to $\mathbb{R}^2$. We'll see more of this in the following chapters.

Plotting points and plugging them into parametric equations can be tedious. Another way to describe a parametric curve is to eliminate parameters.

Example 10.1.2 Eliminating parameters. Rewrite the parametric equations

$$x = 2t^2 - 1$$
$$y = t + 1$$

as a single *Cartesian* equation, i.e., eliminate the parameter t.

Solution. First, solve for t in the second equation to get $t = y - 1$. If we plug this into the first equation, we get an equation in x and y alone:

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

We can see from this equation that the original parametric equations should trace out a rightward opening parabola. Note however that this equation tells us nothing about which portion of the parabola is traced out. □

Example 10.1.3 Parameterizing a Line Segment. Give a set of parametric equations and a corresponding interval that trace out the line segment starting at $(1, -3)$ and ending at $(-2, 4)$.

Solution. First, we'll try to figure out what our equations should look like. We know that x starts at 1 and y starts at -3 on this segment, so a reasonable guess is

$$\begin{aligned} x &= 1 + at \\ y &= -3 + bt \end{aligned}$$

for some unknown constants a and b. To find a and b, we'll specify the interval of values we want t to range over. To make things easier, let's fix t between 0 and 1. Then at $t = 1$ we need to have

$$\begin{aligned} x &= 1 + a = -2 \\ y &= -3 + b = 4 \end{aligned}$$

which gives $a = -3$ and $b = 7$. Therefore the line segment is parameterized by

$$\begin{aligned} &x = 1 - 3t \\ &y = -3 + 7t \\ &0 \le t \le 1. \end{aligned}$$

We can also rewrite these equations in terms of the starting and ending values of x and y. In particular, we have

$$\begin{aligned} &x = 1(1 - t) + (-2)t \\ &y = -3(1 - t) + 4t \\ &0 \le t \le 1. \end{aligned}$$

To see why this works, note that the second term in each equation vanishes at $t = 0$, leaving the starting value for each coordinate. Likewise, the first term in each equation vanishes at $t = 1$, leaving the terminal value for each coordinate. Expanding the equations gives the previous result. Note that this method only works if you select $[0, 1]$ as your interval of values for t. □

Parametric equations are especially useful for tracing movement along a curve that is not the graph of a function, like a circle. In particular, a circle of radius r centered at (h, k) is traced out by the parametric equations

$$\begin{aligned} x &= h + r\cos t \\ y &= k + r\sin t \end{aligned}$$

over the interval $0 \le t \le 2\pi$.

Example 10.1.4 Parametric Equations for Motion on a Circle. Find parametric equations and a corresponding interval of values for t that describe a particle moving along a circle of radius 3 centered at $(3, -2)$ and starting at $(6, -2)$ moving *clockwise* three times around the circle.

Solution. This problem has a lot to unpack, but we know the basic form of our equations will be

$$x = 3 + 3\cos t \quad \text{and} \quad y = -2 + 3\sin t.$$

Since we want to start at $(6, -2)$, we can let $t = 0$ be our starting value. And since we want to move around the circle 3 times, we'll let t vary from 0 to 6π.

Finally, to get clockwise motion we need to replace t with $-t$, getting

$$\begin{aligned} x &= 3 + 3\cos t \\ y &= -2 - 3\sin t \end{aligned}$$

for $0 \leq t \leq 6\pi$. □

Example 10.1.5 Parametric Equations for an Ellipse. Find parametric equations for the ellipse

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

Solution. These equations will look much like the parametric equations for a circle:

$$\begin{aligned} x &= a\cos t \\ y &= b\sin t \end{aligned}$$

for $0 \leq t \leq 2\pi$. □

10.2 Calculus and Parametric Curves

Tangent Lines to Parametric Curves

We have an idea of how to represent curves in the xy-plane using parametric equations. What we'd like to do now is to figure out how to do calculus with parametric equations. To start, we'll find the slope of a parametric curve at some value t. Suppose that

$$\begin{aligned} x &= f(t) \\ y &= g(t). \end{aligned}$$

Then we can find the slope $\frac{dy}{dx}$ using the chain rule:

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

and so $\frac{dy}{dx} = \frac{dy}{dt}/\frac{dx}{dt} = g'(t)/f'(t)$, assuming that $g'(t) \neq 0$.

Example 10.2.1 Cusps on a Cycloid. The parametric equations

$$\begin{aligned} x &= t - \sin t \\ y &= 1 - \cos t \end{aligned}$$

trace out a **cycloid**, also known as the **brachistochrone** or **curve of fastest descent**. Find the values of t for which this curve has a cusp.

Solution. First, note that a curve cannot be differentiable at a cusp. So we can try to find the cusps by looking at where $\frac{dy}{dx}$ fails to exist. This occurs precisely at $t = 2k\pi$ for any integer k. Hence the cycloid should have a cusp at these values of t. □

Area Under Parametric Curves

We can also find areas by using u-substitution and the Fundamental Theorem of Calculus. If $y = F(x)$, then the area under the curve $y = F(x)$ from $x = a$

to $x = b$ is given by $\int_a^b F(x)\,dx$. If we instead have parametric equations

$$\begin{aligned} x &= f(t) \\ y &= g(t), \end{aligned}$$

then $y = F(x) = F(f(t)) = g(t)$ and $dx = f'(t)\,dt$. If the curve goes from $x = a$ to $x = b$ exactly once as t goes from α to β, then the area under the curve is given by

$$\int_\alpha^\beta g(t) f'(t)\,dt.$$

Example 10.2.2 Area of a Circle. Suppose we want to find the area of a circle of radius r. We can do so by parameterizing such a circle with

$$\begin{aligned} x &= r\cos t \\ y &= r\sin t \\ 0 &\le t \le 2\pi. \end{aligned}$$

Then the area of the circle is given by

$$\begin{aligned} 2\int_\pi^0 -r^2 \sin^2 t\,dt &= -2r^2 \int_\pi^0 \frac{1-\cos 2t}{2}\,dt \\ &= -r^2 \left[t - \frac{1}{2}\sin 2t\right]_\pi^0 \\ &= \pi r^2. \end{aligned}$$

One thing to watch out for in the previous calculation is that the limits went from π to 0. This was because of the particular parameterization we chose. □

Arc Length of Parametric Curves

Now we move on to finding lengths of parametric curves. We won't derive the formula here, but it essentially follows from the Pythagorean theorem. We'll see a motivation of the formula later using vector calculus. But for now, suppose we have

$$\begin{aligned} x &= f(t) \\ y &= g(t). \end{aligned}$$

Then the length of the associated parametric curve from $t = \alpha$ to $t = \beta$ is given by

$$\int_\alpha^\beta \sqrt{f'(t)^2 + g'(t)^2}\,dt = \int_\alpha^\beta \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt.$$

Example 10.2.3 Arc Length of a Parametric Curve. Let $x = e^t + e^{-t}$ and $y = 5 - 2t$ for $0 \le t \le 3$. Then the length of this curve is given by

$$\begin{aligned} \int_0^3 \sqrt{(e^t - e^{-t})^2 + 4}\,dt &= \int_0^3 \sqrt{e^{2t} - 2 + e^{-2t} + 4}\,dt \\ &= \int_0^3 \sqrt{e^{2t} + 2 + e^{-2t}}\,dt \\ &= \int_0^3 \sqrt{(e^t + e^{-t})^2}\,dt \end{aligned}$$

$$= \int_0^3 (e^t + e^{-t})\,dt$$
$$= e^3 - e^{-3}.$$

□

SUGGESTED PROBLEMS: 1, 3, 13, 37

10.3 Polar Coordinates

Introducing Polar Coordinates

We typically use **Cartesian** or **rectangular** coordinates to plot points. However, this can lead to issues if we have a graph that isn't rectangular. For example, the circle has a simple geometric description based on its center and radius, but in Cartesian coordinates it can be difficult to work with since every point on the (unit) circle takes the form $(x, \pm\sqrt{1-x^2})$. This square root makes integrals and derivatives complicated. The main problem lies in the fact that we are trying to describe a circle using a rectangular coordinate system. So our goal in this section is to find a more suitable coordinate system for circles.

A circle can be described as the set of all points some fixed distance r from a given point, and we can specify any point on the circle by using an angle θ:

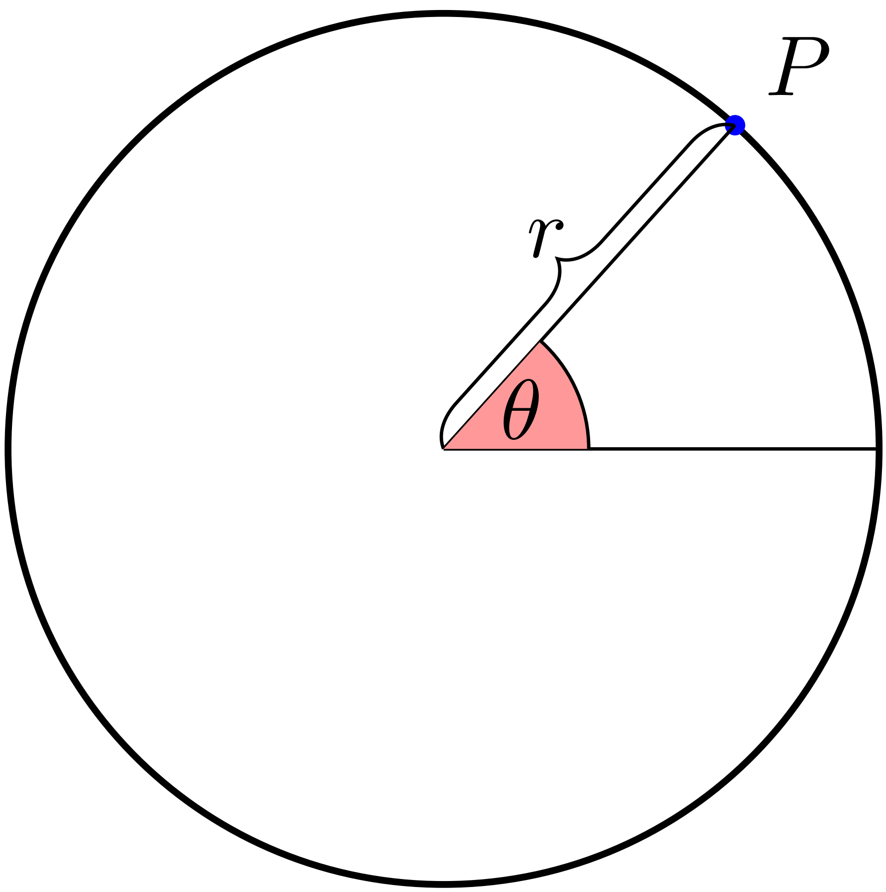


Figure 10.3.1 Using distance and angle to specify a point.

Any point P on the circle can be described solely using the distance r and the direction θ. This leads directly to the idea of polar coordinates. In polar coordinates, the xy-plane is replaced with an $r\theta$-plane, or **polar plane**. Each point in the polar plane has the form (r, θ), where the **radial coordinate** r denotes the distance from the origin, or **pole**, and the **angular coordinate** θ determines the angle the point makes with the horizontal **polar axis**. The polar axis replaces the positive x-axis from Cartesian coordinates:

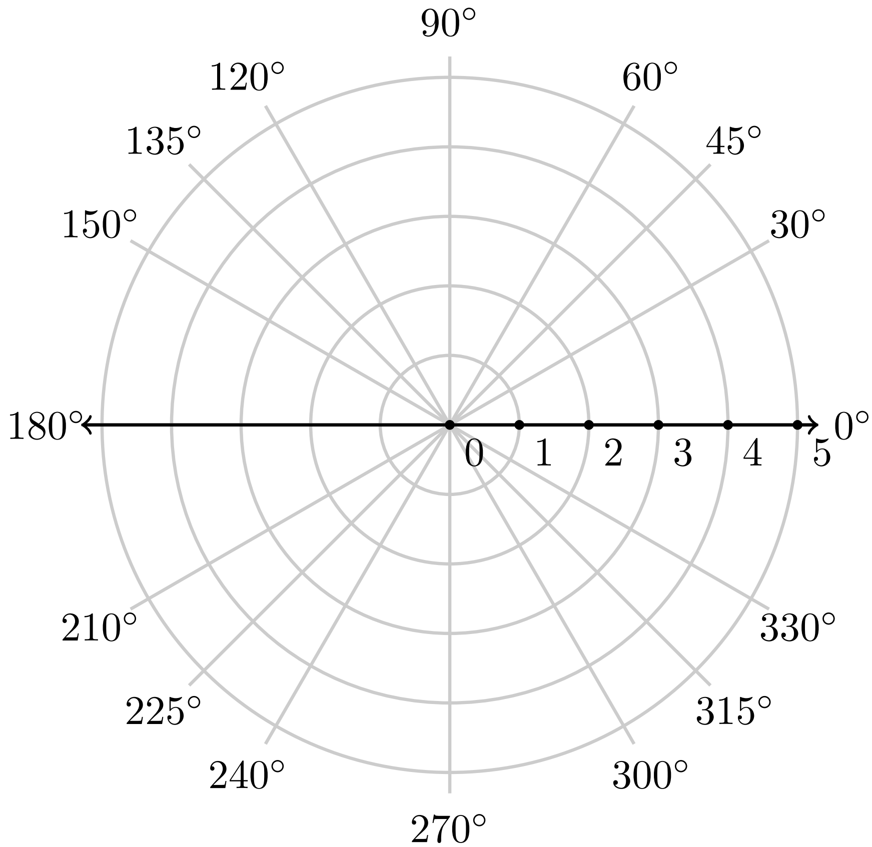


Figure 10.3.2 The polar plane.

In polar coordinates, positive θ correspond to counterclockwise direction, and negative θ correspond to clockwise direction. We also allow for r to be negative: this just means go in the direction opposite of θ.

Example 10.3.3 Plotting polar coordinates. Suppose we want to plot the points $(3, \pi), (-2, \frac{\pi}{4})$ and $(2, \frac{5\pi}{4})$. We can do so in the polar plane by remembering that the first coordinate is distance from the pole, and the second coordinate is direction from the polar axis. If we plot these points, we see that $(-2, \frac{\pi}{4})$ and $(2, \frac{5\pi}{4})$ actually represent the same point. This is typical of polar coordinates: every point has, in general, infinitely many representations. □

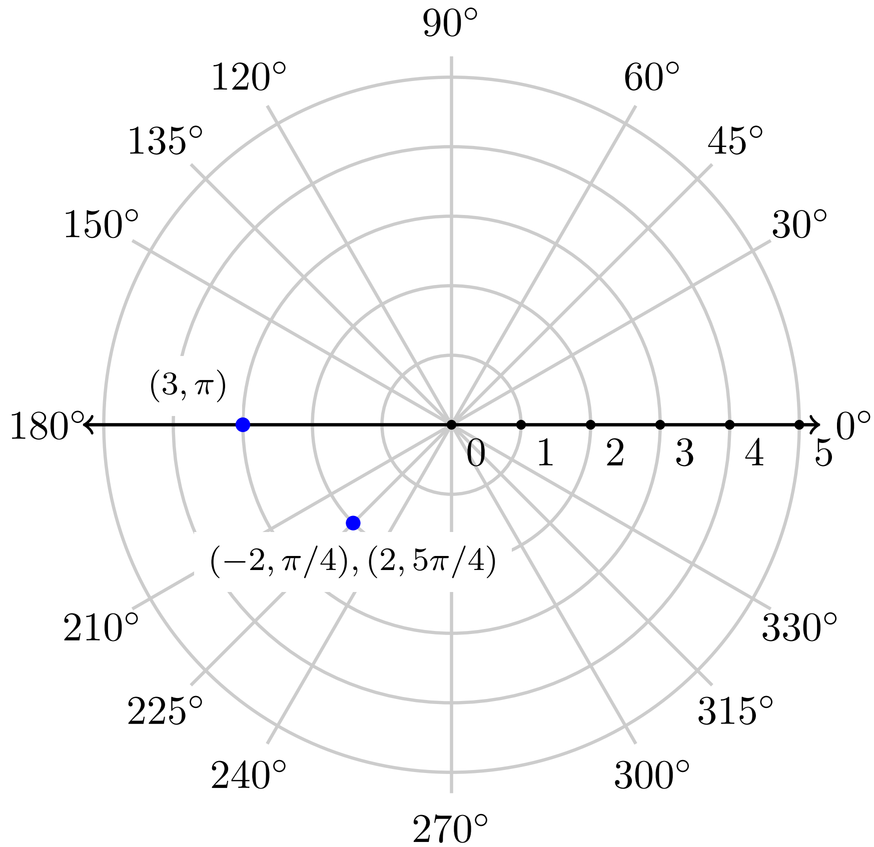


Figure 10.3.4 The plot from Example 10.3.3.

Converting Coordinates

If we want to use polar coordinates, then it'd be helpful to know how to convert between Cartesian (rectangular) coordinates and polar coordinates. The following diagram will help us to make these conversions.

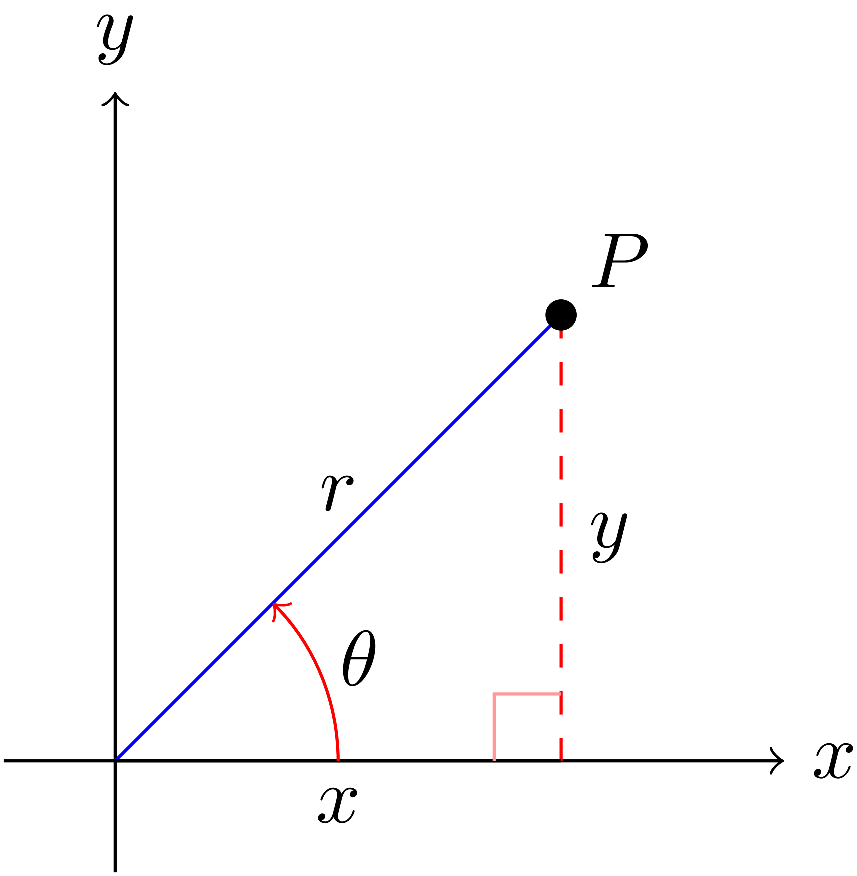


Figure 10.3.5 Converting between Cartesian and polar coordinates.

So in terms of r, θ, we have

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Going in the other direction, the Pythagorean theorem tells us that $r = \sqrt{x^2+y^2}$, while $\tan\theta = \frac{y}{x}$.

Example 10.3.6 Converting polar to Cartesian. Suppose we want to convert the point $(3, \frac{\pi}{3})$ in polar coordinates to Cartesian coordinates (x, y). Then we have

$$x = 3\cos\frac{\pi}{3}$$
$$y = 3\sin\frac{\pi}{3}$$

and so the Cartesian point is $(\frac{3}{2}, \frac{3\sqrt{3}}{2})$. □

Example 10.3.7 Converting Cartesian to polar. Suppose we now want to convert the point $(-\sqrt{3}, 1)$ in Cartesian coordinates to polar coordinates. Then finding r is relatively straightforward:

$$r = \sqrt{3+1} = 2.$$

Finding θ requires a bit more care. We know that θ has to satisfy $\tan\theta = -\frac{1}{\sqrt{3}}$. One choice that makes this work is $\theta = -\frac{\pi}{6}$. However, this is incorrect! Whatever θ is needs to be consistent with the fact that our point lies in the second quadrant. So we'll pick $\theta = \frac{5\pi}{6}$ instead. Hence one way to write the point in polar coordinates is given by $(2, \frac{5\pi}{6})$. □

Example 10.3.8 Converting equations. Consider the polar equation $r\cos\theta = 2\sin\theta\cos\theta$. We can convert this into a Cartesian equation using the above formulas. In particular, we use $r\cos\theta = x$ and $r\sin\theta = y$. So $\cos\theta = \frac{x}{r}, \sin\theta = \frac{y}{r}$ and the equation becomes

$$x = \frac{2xy}{r^2} = \frac{2xy}{x^2+y^2}.$$

□

Polar Curves

Now we move on to graphing polar equations.

Example 10.3.9 Graphs of constants in polar coordinates. In Cartesian coordinates, the graphs of $x = a$ and $y = a$ give horizontal and vertical lines, respectively. In polar coordinates, the graphs of $r = a$ and $\theta = a$ have simple descriptions as well. The graph of $r = a$ is just the set of all points a units away from the pole, so it's just a circle of radius $|a|$ centered at the pole. Likewise, the graph of $\theta = a$ is the set of all points that make an angle of a with the polar axis, so it's just a line through the pole. □

Example 10.3.10 Graphs from Cartesian equations. In some cases it's beneficial to convert a polar equation to a Cartesian equation. The Cartesian equation may have a recognizable form that helps us to identify the corresponding polar graph. For example, suppose we want to describe the graph of $r = 2a\sin\theta$ for some constant a. Then we can convert it to a Cartesian equation, in particular

$$x^2 + y^2 - 2ay = 0.$$

If we complete the square, we get $x^2 + (y-a)^2 = a^2$, which describes a circle centered at $(0, a)$ (in the xy-plane) and with radius $|a|$. In the polar plane, this is a circle of radius $|a|$ centered at $(a, \frac{\pi}{2})$. Similarly, $r = 2a\cos\theta$ describes a circle of radius $|a|$ centered at $(a, 0)$. □

Example 10.3.11 Another circle. Suppose we want to graph $r = -3\cos\theta$. From the previous example, we know that this will be a circle of radius $\frac{3}{2}$ centered at $(-\frac{3}{2}, 0)$. See the following figure. □

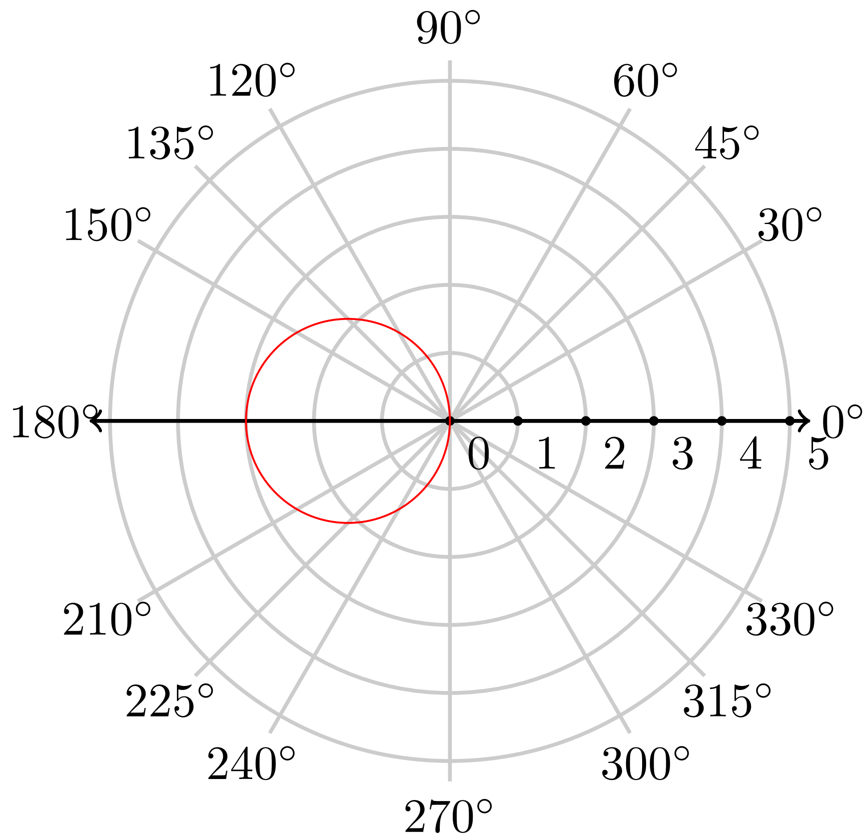


Figure 10.3.12 The circle from Example 10.3.11.

Cartesian to Polar Method

Graphing polar equations can be tricky, because it's easy to miss aspects of the graph unless you're careful. A useful method for graphing polar equations involves treating them as Cartesian equations first to get a better sense of how the graph behaves.

Example 10.3.13 Three leaf rose. Suppose we want to graph $r = \cos 3\theta$. We don't know what this looks like in the polar plane yet, but we have a pretty good idea of how it looks when treated as a Cartesian equation in the $r\theta$-plane (which we view as different from the polar plane!):

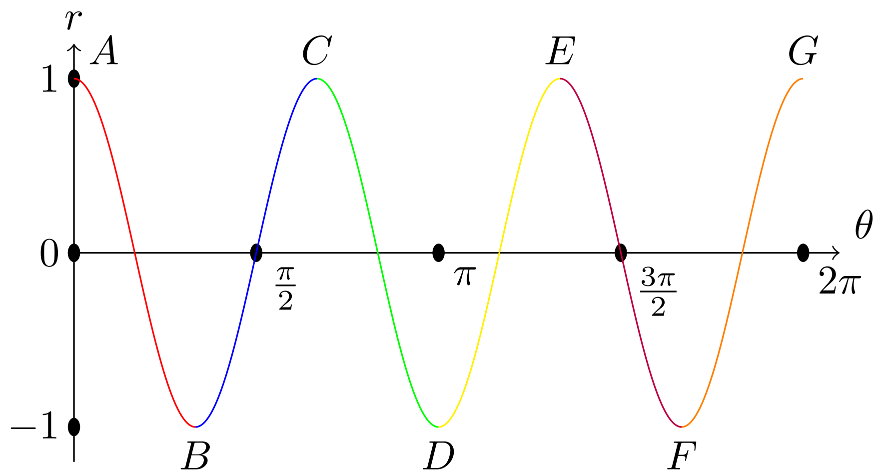


Figure 10.3.14 The graph of $r = \cos 3\theta$ as a Cartesian equation.

So we see that at θ goes from 0 to $\frac{\pi}{3}$, r decreases from 1 to -1, hitting 0 along the way at $\theta = \frac{\pi}{6}$.

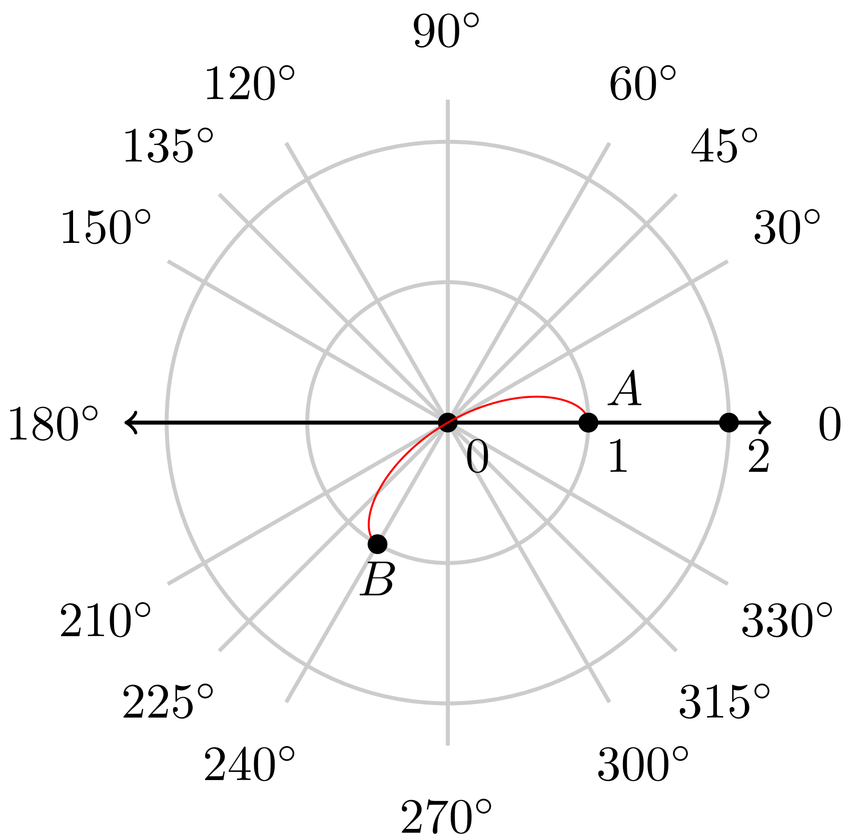


Figure 10.3.15 $\theta = 0$ to $\theta = \frac{\pi}{3}$.

Now, as θ goes from $\frac{\pi}{3}$ to $\frac{2\pi}{3}$ r will go from -1 to 1, hitting 0 at $\theta = \frac{\pi}{2}$:

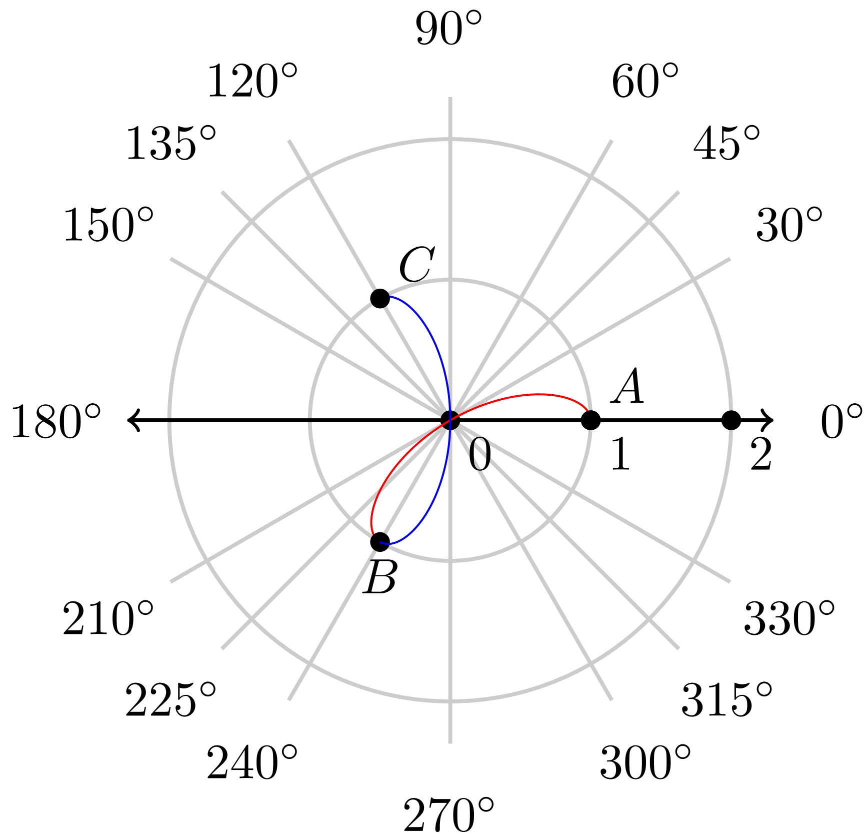


Figure 10.3.16 $\theta = \frac{\pi}{3}$ to $\theta = \frac{2\pi}{3}$.

Continuing in this manner lets us complete the graph:

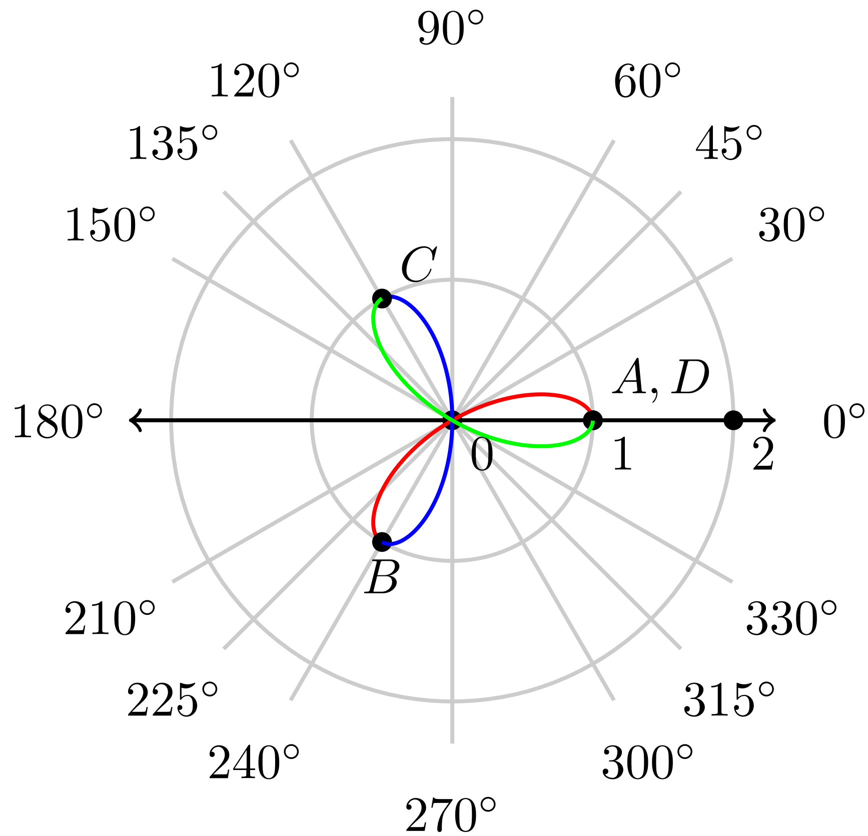


Figure 10.3.17 The graph of $r = \cos 3\theta$ in the polar plane.

□

SUGGESTED PROBLEMS: 1, 3, 11, 23, 37

10.4 Calculus and Polar Coordinates

Areas of Polar Curves

To find the area of a region swept out by a polar curve, we need to reinvent the integral so to speak. In Cartesian coordinates, the integral was based on using rectangles to approximate areas. For polar coordinates, it makes more sense to use circular sectors.

Consider a region R bounded by the polar curve $r = f(\theta)$:

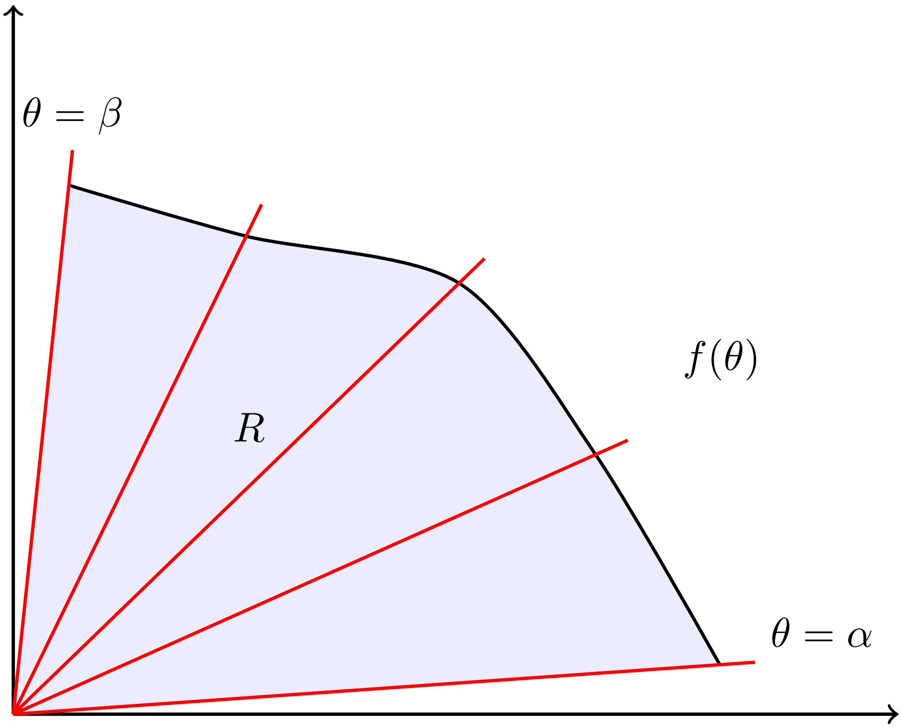


Figure 10.4.1 The polar region $r = f(\theta)$.

To find the area of this region, we'll approximate it using circular sectors. Since we want to find the area between $\theta = \alpha$ and $\theta = \beta$, we'll break the interval $[\alpha, \beta]$ into n subintervals $[\theta_i, \theta_{i+1}]$ of equal length $\Delta\theta$, with

$$\alpha = \theta_0 < \theta_1 < \cdots < \theta_n = \beta.$$

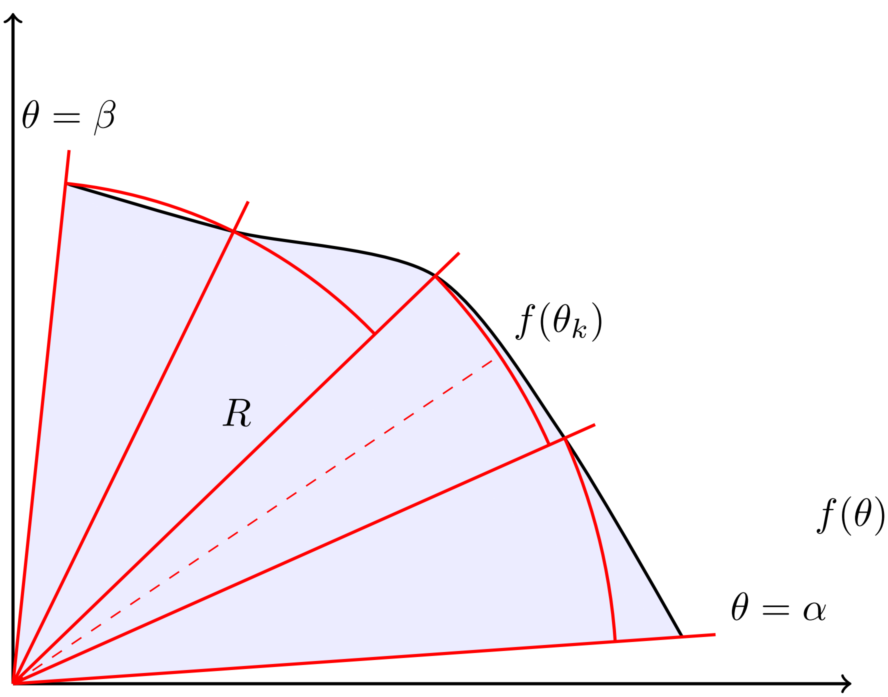


Figure 10.4.2 Approximating the region with circular sectors.

The area of each slice is given by

$$\frac{\text{angle swept out by slice}}{2}(\text{radius})^2.$$

For one of our slices, this takes the form

$$\frac{\Delta\theta}{2}f(\theta_i)^2.$$

Adding these areas up gives us the approximate area of the region. So we can say that

$$\text{area of } R \approx \sum_{i=0}^{n}\frac{1}{2}f(\theta_i)^2\Delta\theta.$$

As $\Delta\theta$ gets smaller, this approximation should become exact. So we can say

$$\text{area of } R = \lim_{\Delta\theta\to 0}\sum_{i=0}^{n}\frac{1}{2}f(\theta_i)^2\Delta\theta.$$

This limit is exactly the definition of an integral! So we have

$$\text{area of } R = \int_{\alpha}^{\beta}\frac{1}{2}f(\theta)^2\,d\theta.$$

We can summarize this in the following theorem.

Theorem 10.4.3 Area in Polar Coordinates. *Suppose that R is a region bounded by the polar curves $r = f(\theta)$ and $r = g(\theta)$ between $\theta = \alpha$ and $\theta = \beta$. If f and g are continuous and $f(\theta) \geq g(\theta)$ for $\alpha \leq \theta \leq \beta$, then the area of R is given by*

$$\int_{\alpha}^{\beta}\frac{1}{2}[f(\theta)^2 - g(\theta)^2]\,d\theta.$$

Example 10.4.4 Area bounded by a cardioid. Consider the cardioid[3] $r = 4 + 4\cos\theta$. The area is given by

$$\int_{\alpha}^{\beta}\frac{1}{2}(4+4\cos\theta)^2\,d\theta = \int_{\alpha}^{\beta}8(1+2\cos\theta+\cos^2\theta)\,d\theta$$

To find this, we'll need to figure out the limits of integration. For polar plots, the best way to do this is by graphing. We'll graph $r = 4+4\cos\theta$ by using the Cartesian to polar methods described in Cartesian to Polar Method. First, we treat this as a Cartesian curve in the $r\theta$-plane and graph it to get

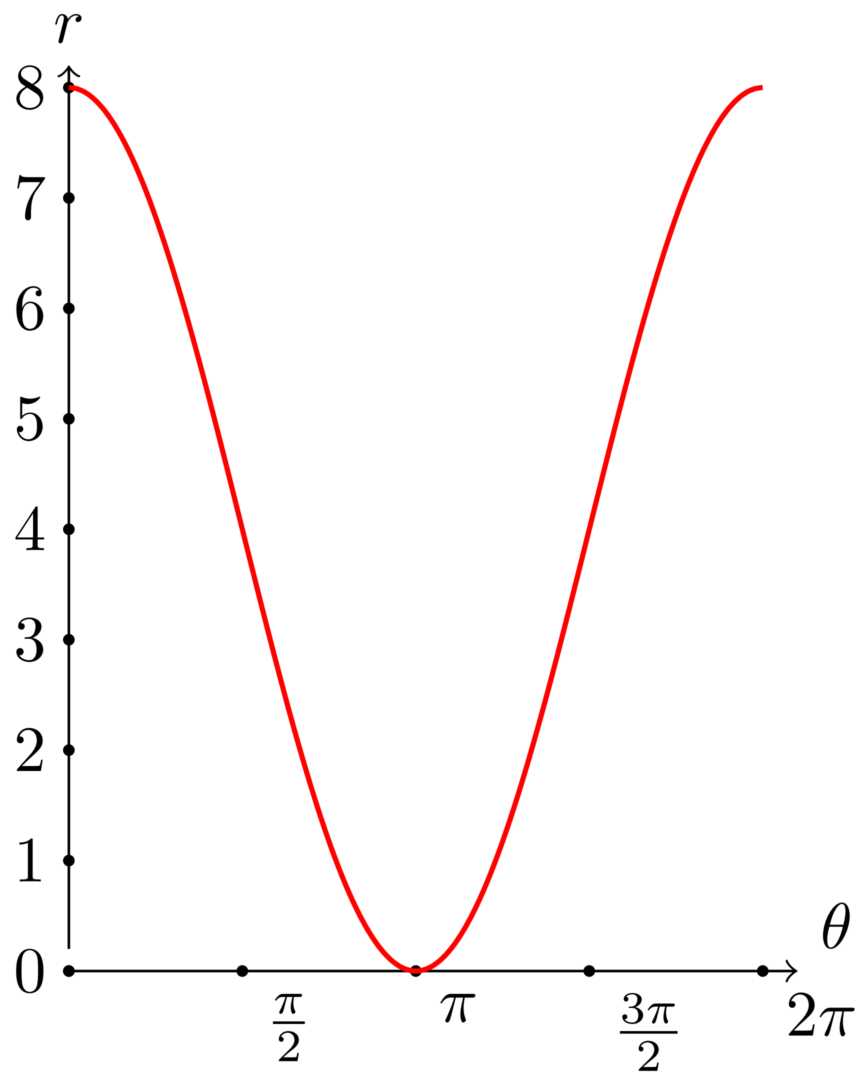


Figure 10.4.5 $r = 4 + 4\cos\theta$ as a Cartesian equation.

Now we can plot the corresponding polar curve:

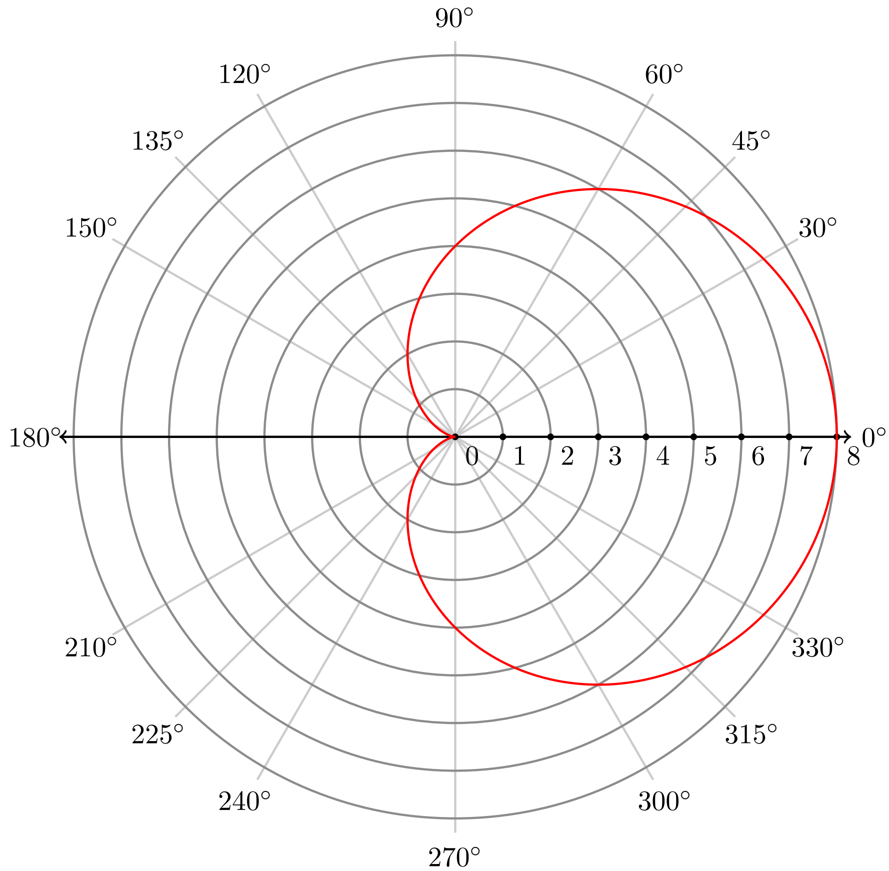


Figure 10.4.6 $r = 4 + 4\cos\theta$ as a polar equation.

The important thing to notice is that we trace out the curve exactly once as θ goes from 0 to 2π, so our limits should be 0 and 2π. Hence the area of this region is given by

$$\begin{aligned}\int_0^{2\pi} 8(1 + 2\cos\theta + \cos^2\theta)\,d\theta &= 8\theta\Big]_0^{2\pi} + 16\sin\theta\Big]_0^{2\pi} + 4\int_0^{2\pi}(1+\cos 2\theta)\,d\theta \\ &= 16\pi + 8\pi \\ &= 24\pi.\end{aligned}$$

□

Example 10.4.7 Area inside a rose and outside a circle. As another example of the use of Theorem 10.4.3, we'll find the area of the region contained by the polar curves $r = 4\cos 2\theta$ and outside the circle $r = 2$. Again, a great way to start this problem is to graph the region under consideration.

[3]The name will become clearer once we graph this.

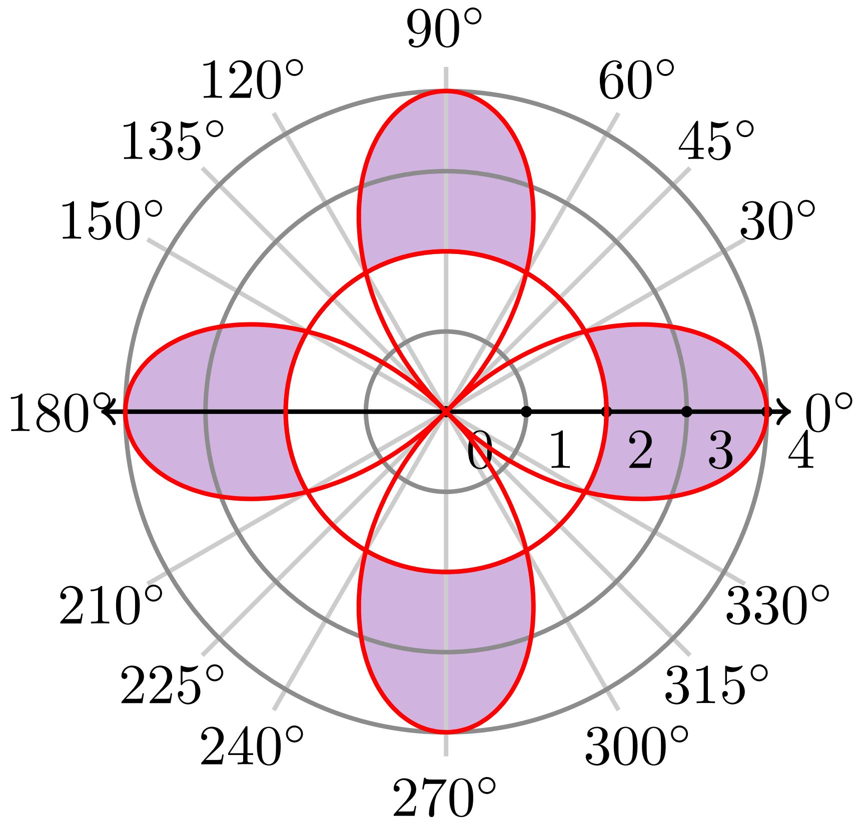


Figure 10.4.8 Region contained between polar curves.

With a little bit of work, we can see that the area of one of these shaded regions is given by

$$\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \frac{1}{2}[(4\cos 2\theta)^2 - 4]\, d\theta$$

and so the area of the entire region is

$$2\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} [(4\cos 2\theta)^2 - 4]\, d\theta.$$

□

Arc Length in Polar Coordinates

Our goal now is to find the length of a polar curve $r = f(\theta)$. From Arc Length of Parametric Curves we know how to find the length of a parametric curve, and a polar curve is really just a parametric curve with parametric equations

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

Using the formula $L = \int_a^b \sqrt{(\frac{dx}{d\theta})^2 + (\frac{dy}{d\theta})^2}\, d\theta$ and simplifying, we get the following result.

Theorem 10.4.9 Arc Length of a Polar Curve. *The length L of a polar* *curve $r = f(\theta)$ from $\theta = a$ to $\theta = b$ is given by*

$$L = \int_a^b \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta.$$

Example 10.4.10 Arc length of a leaf. Let $r = \cos 2\theta$. Then this gives a four-leaf rose in the polar plane. We want to find the length of one leaf of this rose. As usual, it's a good idea to graph this curve first.

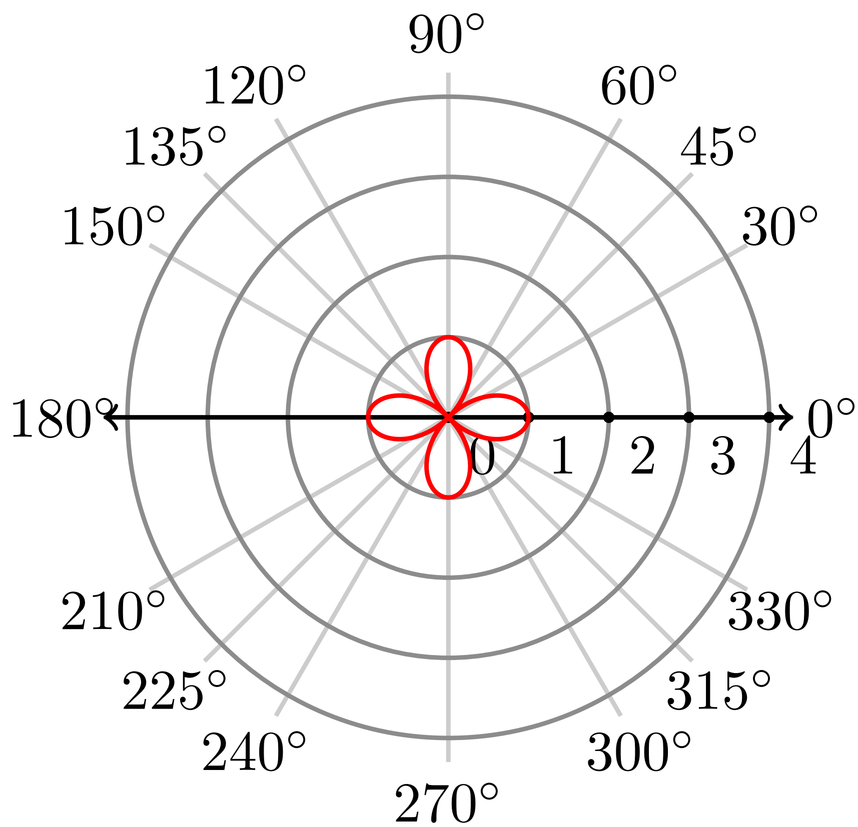


Figure 10.4.11 The four leaf rose $r = \cos 2\theta$.

One leaf of the rose should be captured as θ varies from $-\frac{\pi}{4}$ to $\frac{\pi}{4}$, so using Theorem 10.4.9 we can find the length of this leaf:

$$L = \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \sqrt{\cos^2 2\theta + 4\sin^2 2\theta}\, d\theta = \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \sqrt{1 + 3\sin^2 2\theta}\, d\theta.$$

□

SUGGESTED PROBLEMS: 1, 3, 7, 9, 15, 19, 35

Chapter 11

Vectors and Euclidean Space

Our main goal now is to extend the concepts of calculus into multiple dimensions. We start by considering the three-dimensional space $\mathbb{R}^3$.

11.1 Coordinates in 3-Space

xyz-coordinates

We're used to Cartesian coordinates in the plane: each point in the plane can be represented by an x-coordinate and a y-coordinate. This representation is determined by the **coordinate axes** (the x- and y-axes). There's nothing preventing us from doing the same for three dimensions. We'll just need three coordinate axes: the x-, y- and z-axes. We typically view the xy-plane as horizontal, and the z-axis as vertical, but there's no mathematical preference either way.

Just as the x- and y-axes determined the xy-plane, we can also get the xz- and yz-planes. These planes divide space into eight **octants**. And just as any point in the xy-plane can be represented by measuring along the x-axis and y-axis to get a point (x, y), we can do the same in space to get a point (x, y, z). The set of all points like this is denoted by $\mathbb{R}^3$, and is the **three-dimensional coordinate system**.

Given any point (a, b, c) in $\mathbb{R}^3$, we can find its **projections** onto any of the coordinate axes or coordinate planes without too much trouble. The projection onto a line or plane is the point on that line or plane that is closest to the original point (a, b, c).

Example 11.1.1 Projection. Find the projection of the point $(-1, 3, 4)$ onto the yz-plane.

Solution. Here's how we can find the projection of the point $(-1, 3, 4)$ onto the yz-plane. Note that the yz-plane is just the set of all points with x-coordinate equation to 0, so the projection of $(-1, 3, 4)$ onto the yz-plane is the point $(0, 3, 4)$. □

Example 11.1.2 Equations in space. Sketch $y = 1$ in $\mathbb{R}^3$.

Solution. This is just the set of all points in $\mathbb{R}^3$ of the form $(x, 1, z)$. This forms a **plane** in $\mathbb{R}^3$. □

```
x,z = var('x,z')
parametric_plot3d((x,1,z), (x,-4,4), (z,-4,4))
```

Example 11.1.3 More equations in space. Sketch $x + y = 2$ in $\mathbb{R}^3$.

Solution. We can do so as follows. First, sketch $x + y = 2$ in the xy-plane, which will just be the line $y = 2 - x$. Then the surface in $\mathbb{R}^3$ represented by $x + y = 2$ is actually the surface consisting of all points directly above and directly below the line $y = 2 - x$ in the xy-plane. □

Example 11.1.4 Intersection of a sphere and a plane. Describe the intersection of the sphere given by the equation $(x + 2)^2 + (y - 3)^2 + z^2 = 9$ with the xz-plane.

Solution. We can do this without too much trouble if we remember that the xz-plane is just the set of all points with y-coordinate equal to 0. So the intersection of this sphere with the xz-plane traces out the curve $(x + 2)^2 + (0 - 3)^2 + z^2 = 9$ in the xz-plane, which is just the point $(-2, 0, 0)$. □

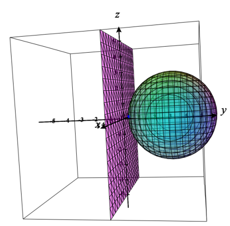


Figure 11.1.5 Intersection of the sphere $(x + 2)^2 + (y - 3)^2 + z^2 = 9$ with the xz-plane.

The Distance Formula

Recall that the distance between two points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ in $\mathbb{R}^2$ (the xy-plane) is given by

$$|P_1P_2| = d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

This is proved using the Pythagorean theorem. We can do the same exact thing in $\mathbb{R}^3$!

Theorem 11.1.6 Distances in Space. *Let $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ be two points in $\mathbb{R}^3$. Then the distance between these two points, $|P_1P_2| = d(P_1, P_2)$, is given by*

$$|P_1P_2| = d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

Example 11.1.7 Computing distances. Theorem 11.1.6 lets us find the distance between the points $P(1, 4, -2)$ and $Q(-1, 0, 2)$ as follows:

$$d(P, Q) = \sqrt{4 + 16 + 16} = 6.$$

□

One important use of the distance formula in $\mathbb{R}^3$ is that it lets us find equations of spheres. The equation of a sphere of radius r and center $C(h, k, l)$ is given by

$$\sqrt{(x - h)^2 + (y - k)^2 + (z - l)^2} = r,$$

which is more commonly written as

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

Example 11.1.8 Equation of a sphere. The equation $2x^2 + 2y^2 + 2z^2 = 8x - 24z + 4$ represents a sphere in $\mathbb{R}^3$. To see how, we can rearrange the equation and complete the square to get

$$\begin{aligned} 2x^2 + 2y^2 + 2z^2 - 8x + 24z = 4 &\Rightarrow x^2 + y^2 + z^2 - 4x + 12z = 2 \\ &\Rightarrow (x-2)^2 + y^2 + (z-6)^2 = 2 + 4 + 36 \\ &\Rightarrow (x-2)^2 + y^2 + (z-6)^2 = 42. \end{aligned}$$

So this equation describes a sphere of radius $\sqrt{42}$ centered at $(2, 0, 6)$. □

Example 11.1.9 Spherical shells. We can also use inequalities to describe regions in addition to equalities. For example, $3^2 \leq x^2 + y^2 + z^2 \leq 4^2$ describes the region contained between the sphere of radius 3 and the sphere of radius 4, both centered at the origin. □

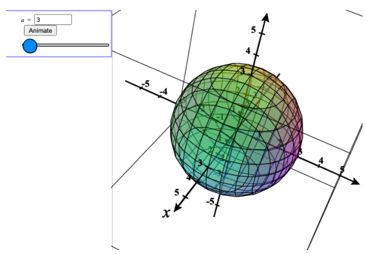


Figure 11.1.10 The spherical shells $x^2 + y^2 + z^2 = a^2, 3 \leq a \leq 4$.

11.2 Vectors

One of our goals in this chapter is to adequately describe motion in space. A useful way to do this uses the concept of **vector**, which we think of as a quantity that has both direction and magnitude/length. A simple example would be velocity: velocity in space has a direction and also a magnitude (speed). We will typically denote vectors by using boldface letters such as $\mathbf{x}$ or letters with a bar overhead such as $\bar{x}$. We represent vectors as arrows with an **initial point** and a **terminal point**:

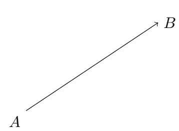


Figure 11.2.1 A vector.

We say that two vectors $\mathbf{x}, \mathbf{y}$ are **equivalent**, or **equal**, if they have the same magnitude and direction. We write this as $\mathbf{x} = \mathbf{y}$.

Addition and Scalar Multiplication

Given two vectors $\mathbf{x} = \overrightarrow{AB}, \mathbf{y} = \overrightarrow{BC}$, we can add them to get the new vector

$$\mathbf{x} + \mathbf{y} = \overrightarrow{AC}.$$

So the vector $\mathbf{x} + \mathbf{y}$ is obtained by moving the tail of $\mathbf{y}$ to the tip of $\mathbf{x}$ and then drawing a vector from the tail of $\mathbf{x}$ to the tip of $\mathbf{y}$.

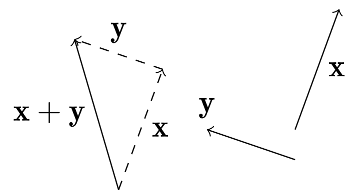


Figure 11.2.2 Vector addition.

The sum $\mathbf{x}+\mathbf{y}$ of vectors $\mathbf{x},\mathbf{y}$ can be computed using either the **triangle law**, illustrated above in Figure 11.2.2, or the similar **parallelogram law**. We can also scale vectors using **scalar multiplication**: if α is a scalar[4] then $\alpha\mathbf{x}$ is defined to be the vector that has the same direction as $\mathbf{x}$ if $\alpha > 0$ and the opposite direction if $\alpha < 0$, but the magnitude is rescaled by the factor α.

Example 11.2.3 Vector subtraction. Using the previous graph, we can compute $2\mathbf{x}-\mathbf{y}$. We just need to scale the vectors properly and then add $2\mathbf{x}$ to $-\mathbf{y}$. □

Vector Components

Although it isn't too hard to add and scale vectors visually, it'll be beneficial to do the same algebraically. We can do so by breaking a vector down into its **components**. Consider a vector in $\mathbf{x}$ in $\mathbb{R}^2$, and suppose if we move it to the origin then the tip of the vector is at the point $(3,-1)$. Then the components of $\mathbf{x}$ are 3 and -1, and we write

$$\mathbf{x} = \langle 3,-1\rangle .$$

Note the use of brackets here, since technically we are saying that the vector $\mathbf{x}$ is distinct from the point $(3,-1)$, even though they are closely related. We say that the **position vector** of a point is the vector whose components are the same as the coordinates of the point. Geometrically, the position vector of a point P is the vector with its tail at the origin and its tip at P.

So any vector in $\mathbb{R}^2$ can be represented using components by $\mathbf{x} = \langle x_1, x_2\rangle$. Similarly, any vector in $\mathbb{R}^3$ can be represented as $\mathbf{x} = \langle x_1, x_2, x_3\rangle$. Once you represent a vector in component form, addition and scalar multiplication is straightforward.

Example 11.2.4 Vector addition with components. Let $\mathbf{a} = \langle -2,1,3\rangle$ and $\mathbf{b} = \langle 0,2,1\rangle$. Then

$$\mathbf{b}-3\mathbf{a} = \langle 6,-1,-8\rangle .$$

□

Finding magnitudes of vectors can also be done by applying the distance formula from Section 11.1 to the components of the vectors. The magnitude of a vector $\mathbf{x}$ is denoted by $|\mathbf{x}|$ or $\|\mathbf{x}\|$. For example, the magnitude of $\mathbf{b}-3\mathbf{a}$ from the previous example is

$$\|\mathbf{b}-3\mathbf{a}\| = \sqrt{36+1+64} = \sqrt{101}.$$

Given a vector $\mathbf{x} = \langle x_1, x_2, x_3\rangle$,

$$\|\mathbf{x}\| = \sqrt{x_1^2+x_2^2+x_3^2}.$$

Example 11.2.5 A vector equation. We can use vectors to describe curves and surfaces. For example, let $\mathbf{r} = \langle x,y,z\rangle$ and $\mathbf{r}_0 = \langle a,b,c\rangle$. Let $r \geq 0$. Then

[4] In other words, a number.

the set of all points (x, y, z) that satisfy the equation

$$\|\mathbf{r} - \mathbf{r}_0\| = r$$

has a very nice description: it's just the sphere of radius r centered at (a, b, c).
□

Example 11.2.6 Finding components of vectors. Consider the points $A(4, 0, -2)$ and $B(4, 2, 1)$. We want to find the components of the vector $\overrightarrow{AB}$. We can do this by translating $\overrightarrow{AB}$ to the origin, which is done by subtracting from each coordinate of B the corresponding coordinate of A. So the vector $\overrightarrow{AB}$ is given by

$$\overrightarrow{AB} = \langle 4 - 4, 2 - 0, 1 + 2 \rangle = \langle 0, 2, 3 \rangle .$$

□

In general, given $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$, the vector $\mathbf{x} = \overrightarrow{AB}$ is given by

$$\mathbf{x} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle .$$

Theorem 11.2.7 Properties of Vector Addition and Scalar Multiplication. *Let* $\mathbf{x}, \mathbf{y}, \mathbf{z}$ *be vectors and let* α, β *be scalars. Then the following are true:*

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.
3. $\mathbf{x} + \mathbf{0} = \mathbf{x}$.
4. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
5. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$.
6. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{y}$.
7. $(\alpha\beta)\mathbf{x} = \alpha(\beta)\mathbf{x}$.
8. $1\mathbf{x} = \mathbf{x}$.

Basis Vectors and Unit Vectors

Every vector in $\mathbb{R}^3$ can be written using three components: $\mathbf{x} = \langle x_1, x_2, x_3 \rangle$. Each component corresponds to a coordinate axis, and we can rewrite $\mathbf{x}$ as a **linear combination** of three different vectors, with each vector corresponding to a coordinate axis:

$$\mathbf{x} = x_1 \langle 1, 0, 0 \rangle + x_2 \langle 0, 1, 0 \rangle + x_3 \langle 0, 0, 1 \rangle .$$

These vectors are important enough that we'll give them a name: the **standard basis vectors**.

Definition 11.2.8 Standard Basis Vectors. The standard basis for $\mathbb{R}^3$ is the set $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, where

$$\mathbf{i} = \langle 1, 0, 0 \rangle , \quad \mathbf{j} = \langle 0, 1, 0 \rangle \quad \text{and} \quad \mathbf{k} = \langle 0, 0, 1 \rangle .$$

◊

As we've seen, every vector $\mathbf{x}$ in $\mathbb{R}^3$ can be expressed using only these three vectors. The standard basis has two important properties: it is perpendicular (also called **orthogonal**) and every vector in the collection has magnitude 1. In other words,

$$\|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1.$$

These vectors are essentially designed to capture the "coordinate directions", and are plotted below.

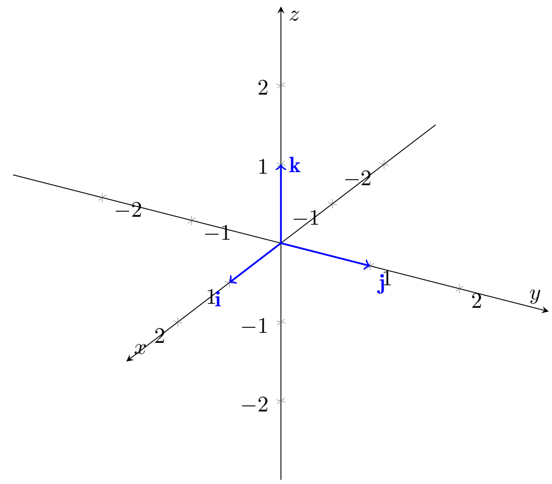


Figure 11.2.9 The standard basis.

This also leads us to our next definition.

Definition 11.2.10 Unit Vectors. A vector $\mathbf{x}$ is a unit vector if $\|\mathbf{x}\| = 1$. ♢

Unit vectors are useful if we just need to indicate a direction, and we don't care about magnitude. Every nonzero vector can be rescaled to a unit vector: just divide the vector by its norm.

Example 11.2.11 Direction from one point to another. Consider the points $A(1, 2, -1)$ and $B(0, -3, 2)$. Then we can find the unit vector indicating the direction from A to B. First, set

$$\mathbf{x} = \overrightarrow{AB} = \langle -1, -5, 3 \rangle .$$

Then the unit vector that gives the direction from A to B is given by

$$\begin{aligned} \mathbf{u} &= \frac{\mathbf{x}}{\|\mathbf{x}\|} \\ &= \frac{\langle -1, -5, 3 \rangle}{\sqrt{1 + 25 + 9}} \\ &= \left\langle -\frac{1}{\sqrt{37}}, -\frac{5}{\sqrt{37}}, \frac{3}{\sqrt{37}} \right\rangle . \end{aligned}$$

□

Example 11.2.12 A vector equation for the unit sphere. Using the concept of a unit vector, we can very easily describe the unit sphere[5] using a vector equation. If we set $\mathbf{r} = (x, y, z)$, then the unit sphere is just the set of all solutions of

$$\|\mathbf{r}\| = 1.$$

□

Applications

Many physical quantities have both a direction and a magnitude, like velocity, acceleration and forces. Vectors are ideally suited to measure these quantities.

[5] The sphere of radius 1 centered at the origin.

Example 11.2.13 Weight of a chain. A still chain is fixed to two ends of a level divide. The tension of the chain at each fixed end can be represented by vectors pointing away from the chain. Call these tension forces $\mathbf{T}_1$ and $\mathbf{T}_2$. Suppose we know that each force vector makes an angle of 45° with the ground on either side of the chain's fixed ends, and that the magnitude of each tension is 43 N. Then we can use vector addition to find the weight of the chain.

Let $\mathbf{w}$ denote the weight of the chain considered as a vector (so that it's pointing down). Since the chain is still, its **resultant**[6] must be $\mathbf{0}$. So we can say that

$$\mathbf{T}_1 + \mathbf{T}_2 + \mathbf{w} = \mathbf{0}$$

or in other words

$$\mathbf{w} = -\mathbf{T}_1 - \mathbf{T}_2.$$

What we need to find is $\|\mathbf{w}\|$, which we can do without too much trouble if we can rewrite the tension vectors in component form. In fact, we have

$$\begin{aligned}\mathbf{T}_1 &= \langle -\|\mathbf{T}_1\|\cos 45°, \|\mathbf{T}_1\|\sin 45°\rangle\\ \mathbf{T}_2 &= \langle \|\mathbf{T}_2\|\cos 45°, \|\mathbf{T}_2\|\sin 45°\rangle\end{aligned}$$

So it follows that

$$\begin{aligned}\|\mathbf{w}\| &= \|-\mathbf{T}_1 - \mathbf{T}_2\|\\ &= \left\|\left\langle 0, -2\cdot 43\frac{\sqrt{2}}{2}\right\rangle\right\|\\ &= 43\sqrt{2}\end{aligned}$$

Therefore the chain weighs $43\sqrt{2}$N. □

SUGGESTED PROBLEMS: 1--17 odd, 23, 25, 27

11.3 The Dot Product

Definition and Properties of the Dot Product

Suppose we're given two vectors. What we'd like to do is to come up with a measure of how much these vectors overlap. Such a measure may be useful for determining if forces are too close together, for example. So consider two vectors $\mathbf{x} = \langle x_1, y_1, z_1\rangle$ and $\mathbf{y} = \langle x_2, y_2, z_2\rangle$ as in the following diagram.

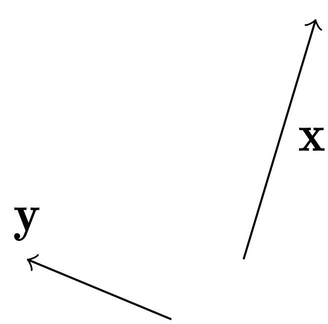


Figure 11.3.1 Vector overlap.

One way we can measure how much $\mathbf{x}$ and $\mathbf{y}$ overlap is to find $\|\mathbf{x}+\mathbf{y}\|$, or

[6]The sum of all forces acting on the chain.

equivalently $\|\mathbf{x}+\mathbf{y}\|^2$, since this is larger if $\mathbf{x}$ and $\mathbf{y}$ are more closely aligned.[7] From Vector Components, we know that

$$\|\mathbf{x}+\mathbf{y}\|^2 = (x_1+x_2)^2 + (y_1+y_2)^2 + (z_1+z_2)^2.$$

This simplifies out to

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= (x_1+x_2)^2 + (y_1+y_2)^2 + (z_1+z_2)^2\\ &= (x_1^2+y_1^2+z_1^2) + 2(x_1x_2+y_1y_2+z_1z_2) + (x_2^2+y_2^2+z_2^2)\\ &= \|\mathbf{x}\|^2 + 2(x_1x_2+y_1y_2+z_1z_2) + \|\mathbf{y}\|^2.\end{aligned}$$

The only part of this that could possibly depend on how closely $\mathbf{x}$ and $\mathbf{y}$ overlap is the middle term $2(x_1x_2+y_1y_2+z_1z_2)$. So we'll (optimistically... for now) take what's inside the parentheses and use it as the measure we're looking for.

Definition 11.3.2 The Dot Product. Let $\mathbf{x} = \langle x_1, y_1, z_1\rangle$ and $\mathbf{y} = \langle x_2, y_2, z_2\rangle$. The **dot product** of $\mathbf{x}$ with $\mathbf{y}$, denoted $\mathbf{x}\cdot\mathbf{y}$, is given by

$$\mathbf{x}\cdot\mathbf{y} = x_1x_2 + y_1y_2 + z_1z_2.$$

◊

The dot product is also sometimes called the **scalar product** (since it always produces a scalar) or the **inner product** in other settings.

Example 11.3.3 Computing dot products. Let $\mathbf{x} = \langle 1, 3, -1\rangle$, $\mathbf{y} = \langle -3, 1, 0\rangle$ and $\mathbf{z} = \langle 1, 1, 1\rangle$. Then

$$\mathbf{x}\cdot\mathbf{z} = 1+3-1 = 3 \quad\text{and}\quad \mathbf{x}\cdot\mathbf{y} = -3+3+0 = 0.$$

It also doesn't matter what order we take the vectors in for the dot product: we get the same answer regardless. However, it does matter that we only use two vectors. The dot product takes two vectors and gives a scalar. In other words, *it is impossible to take the dot product of more than two vectors at a time!* So quantities such as $\mathbf{x}\cdot\mathbf{y}\cdot\mathbf{z}$ are not meaningful. □

Theorem 11.3.4 Properties of the Dot Product. *Let* $\mathbf{x}, \mathbf{y}$ *and* $\mathbf{z}$ *be vectors, and let* α *be a scalar. Then the following properties hold:*

1. $\mathbf{x}\cdot\mathbf{x} = \|\mathbf{x}\|^2$.
2. $\mathbf{x}\cdot\mathbf{y} = \mathbf{y}\cdot\mathbf{x}$.
3. $\mathbf{x}\cdot(\mathbf{y}+\mathbf{z}) = \mathbf{x}\cdot\mathbf{y} + \mathbf{x}\cdot\mathbf{z}$.
4. $(\alpha\mathbf{x})\cdot\mathbf{y} = \alpha(\mathbf{x}\cdot\mathbf{y}) = \mathbf{x}\cdot(\alpha\mathbf{y})$.
5. $\mathbf{0}\cdot\mathbf{x} = 0$.

Our goal was to define a measure for how much two given vectors align, or are correlated. The following result tells us that the dot product is actually a reasonable measure of this.

Theorem 11.3.5 Geometry of the Dot Product. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *denote vectors, and let* $0 \le \theta \le \pi$ *denote the angle between these vectors if the tails of both are moved to the origin. Then*

$$\mathbf{a}\cdot\mathbf{b} = \|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta.$$

Proof. Place both vectors $\mathbf{a}$ and $\mathbf{b}$ at the origin, and then draw the vector $\mathbf{a}-\mathbf{b}$ from the tip of $\mathbf{b}$ to the tip of $\mathbf{a}$, like so:

[7] We are carefully ignoring the case where $\mathbf{x}$ and $\mathbf{y}$ point in opposite directions...

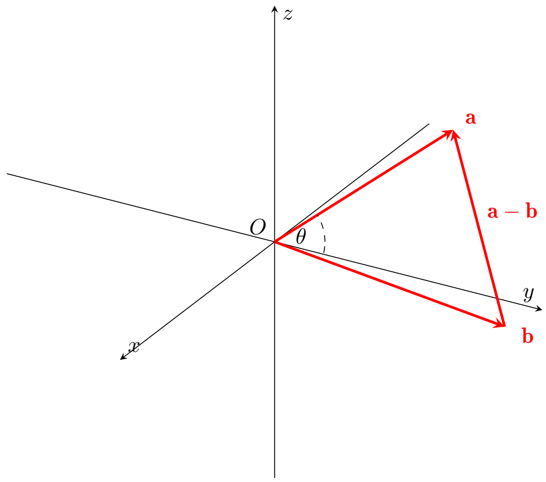


Figure 11.3.6 Geometry of the dot product.

Then $\mathbf{a}, \mathbf{b}$ and $\mathbf{a}-\mathbf{b}$ form a triangle. We want to relate the dot product $\mathbf{a}\cdot\mathbf{b}$ with the angle θ, so we'll start by using the Law of Cosines to get an equation involving θ. The Law of Cosines states that

$$\|\mathbf{a}-\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\,\|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta. \tag{III.1}$$

Here's where the dot product comes in. Each squared magnitude can be rewritten as a dot product using Theorem 11.3.4, so in particular we have

$$\|\mathbf{a}-\mathbf{b}\|^2 = (\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b}) = \|\mathbf{a}\|^2 - 2(\mathbf{a}\cdot\mathbf{b}) + \|\mathbf{b}\|^2 .$$

Plugging this into (III.1) gives us the following:

$$\|\mathbf{a}\|^2 - 2(\mathbf{a}\cdot\mathbf{b}) + \|\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\,\|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta$$

Which finally simplifies down to

$$\mathbf{a}\cdot\mathbf{b} = \|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta.$$

■

Remark 11.3.7 It's usually easier to use Definition 11.3.2 to compute dot products. However, Theorem 11.3.5 gives us extremely useful geometric information about the dot product. For example, we get the following result very quickly: two vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular if and only if $\mathbf{a}\cdot\mathbf{b} = 0$.

Example 11.3.8 Checking orthogonality using the dot product. Let $\mathbf{a} = \langle 0, 1, -3\rangle$ and $\mathbf{b} = \langle 1, 1, 2\rangle$. Then we can say right away that these vectors are *not* orthogonal to each other since $\mathbf{a}\cdot\mathbf{b} = -5 \neq 0$. □

Example 11.3.9 Finding angles between lines. Consider the lines $x + 2y = 7$ and $5x - y = 2$ in $\mathbb{R}^2$, plotted below:

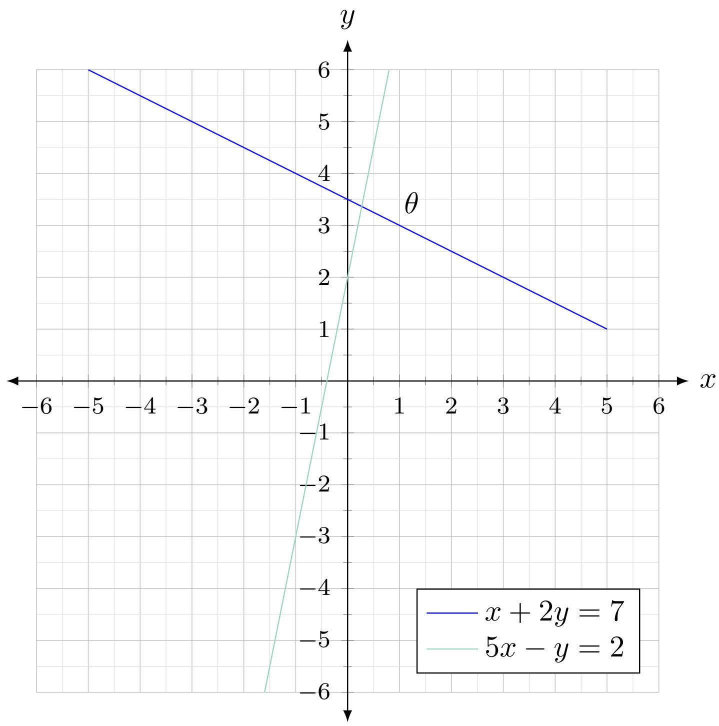


Figure 11.3.10 Angle between lines.

Suppose we want to find the acute angle θ these lines make with each other. We can do so by finding vectors $\mathbf{a}$ and $\mathbf{b}$ that point in the same directions as these lines. We'll start by finding the vector $\mathbf{a}$ which points in the same direction as the line $x + 2y = 7$. To do so, we need two points on this line: $A_1 = (7, 0)$ and $A_2 = (5, 1)$. So we can take $\mathbf{a}$ to be

$$\mathbf{a} = \overrightarrow{A_1 A_2} = \langle -2, 1 \rangle .$$

Similarly, since $B_1 = (0, -2)$ and $B_2 = (1, 3)$ both lie on $5x - y = 2$, we can take

$$\mathbf{b} = \overrightarrow{B_1 B_2} = \langle 1, 5 \rangle .$$

Then by Theorem 11.3.5 we know that

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \, \|\mathbf{b}\|} = \frac{3}{\sqrt{5}\sqrt{26}} = \frac{3}{\sqrt{130}}.$$

So the acute angle between these two lines is given by

$$\theta = \cos^{-1} \frac{3}{\sqrt{130}}.$$

□

Projections

Consider two vectors $\mathbf{a}$ and $\mathbf{b}$ arranged as follows:

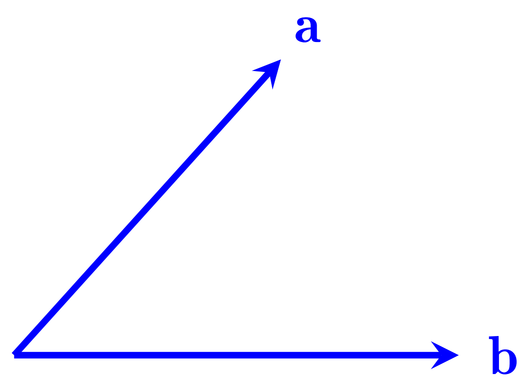


Figure 11.3.11 The vectors **a** and **b**.

Now draw a line from the tip of **a** to the point on **b** that is closest to the tip of **a**; such a line must be orthogonal to **b**. This point on **b** defines a new vector that we call the **projection of a onto b**; we denote this vector by $\operatorname{proj}_{\mathbf{b}} \mathbf{a}$.

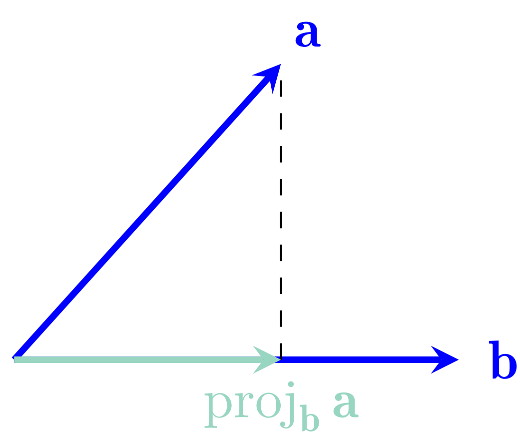


Figure 11.3.12 The projection of **a** onto **b**.

The projection $\operatorname{proj}_{\mathbf{b}} \mathbf{a}$ can be thought of as the vector parallel to **b** that best approximates **a**. What we'd like to do now is to actually find a formula for this projection. If we let θ denote the acute angle between the vectors **a** and **b**, then

$$\|\operatorname{proj}_{\mathbf{b}} \mathbf{a}\| = \|\mathbf{a}\| \cos\theta.$$

Since the projection must also be parallel to **b**, then this is enough information to state exactly what $\operatorname{proj}_{\mathbf{b}} \mathbf{a}$ should be: $\operatorname{proj}_{\mathbf{b}} \mathbf{a} = (\|\mathbf{a}\| \cos\theta) \frac{\mathbf{b}}{\|\mathbf{b}\|}$. We can simplify this formula somewhat by making use of Theorem 11.3.5.

Theorem 11.3.13 Vector Projection Formulas. *Let* **a** *and* **b** *denote vectors in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$*. Then the projection of* **a** *onto* **b** *is given by*

$$\operatorname{proj}_{\mathbf{b}} \mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}} \mathbf{b}.$$

Example 11.3.14 Finding vector projections. Let $\mathbf{a} = \langle 1, 4 \rangle$ and $\mathbf{b} = \langle 2, 3 \rangle$. Then by Theorem 11.3.13 the projection of **a** onto **b** is given by

$$\operatorname{proj}_{\mathbf{b}} \mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}} \mathbf{b} = \frac{14}{13} \langle 2, 3 \rangle.$$

□

Another example of vector projection is computing work done by a force. In particular, suppose we have a force $\mathbf{F}$ and some displacement vector $\mathbf{D}$. We define the work done by $\mathbf{F}$ along $\mathbf{D}$ to be the product of the component of $\mathbf{F}$ along $\mathbf{D}$ times the distance moved. If we let θ denote the acute angle between $\mathbf{F}$ and $\mathbf{D}$, then the work W done is given by

$$W = (\|\mathbf{F}\| \cos\theta) \|\mathbf{D}\|,$$

which is exactly equal to $\mathbf{F} \cdot \mathbf{D}$ by Theorem 11.3.5.

Example 11.3.15 Finding work done by a force. Suppose a force $\mathbf{F} = 2\mathbf{i} - 3\mathbf{j} + \mathbf{k}$ moves a particle from the point $(2, 3, -1)$ to the point $(1, 0, -3)$, where the force is measured in newtons and the distance in meters. We want to find the work done by this force on the particle. To do so, we need the displacement $\mathbf{D}$:

$$\mathbf{D} = \langle 1-2, 0-3, -3-(-1)\rangle = \langle -1, -3, -2\rangle.$$

So the work done is

$$W = \mathbf{F} \cdot \mathbf{D} = -2 + 9 - 2 = 5$$

joules. □

SUGGESTED PROBLEMS: 1--19 odd, 35, 39

11.4 The Cross Product

The dot product, and in particular Theorem 11.3.5, gives us a good way to tell if two vectors are perpendicular. However, it says nothing about how to *construct* perpendicular vectors. The next vector operation, the **cross product**, is the tool we'll use for that goal.

Definition and Properties of the Cross Product

Definition 11.4.1 The Cross Product. Let $\mathbf{x} = \langle x_1, y_1, z_1\rangle$ and $\mathbf{y} = \langle x_2, y_2, z_2\rangle$. Then the cross product of $\mathbf{x}$ with $\mathbf{y}$ is the new vector $\mathbf{x} \times \mathbf{y}$ given by

$$\mathbf{x} \times \mathbf{y} = \langle y_1 z_2 - z_1 y_2, z_1 x_2 - x_1 z_2, x_1 y_2 - y_1 x_2\rangle.$$

◊

Example 11.4.2 Cross product of basis vectors. Let's start by computing $\mathbf{i} \times \mathbf{k}$ using the definition. If we do so, we have

$$\mathbf{i} \times \mathbf{k} = \langle 0, -1, 0\rangle = -\mathbf{j}.$$

On the other hand, we also have $\mathbf{k} \times \mathbf{i} = \mathbf{j}$. This points out the very important fact that *order matters for cross products.* □

This formula is a lot to remember, so it's beneficial to find another way to express it. One way is by using **determinants**. In particular, if $\mathbf{x} = \langle x_1, y_1, z_1\rangle$ and $\mathbf{y} = \langle x_2, y_2, z_2\rangle$, then

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} \tag{III.2}$$

Example 11.4.3 Another cross product. (III.2) is useful to use if you're dealing with vectors that aren't as simple as the basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$. For example, let $\mathbf{a} = \langle -1, 2, 0\rangle$ and $\mathbf{b} = \langle 2, 4, 1\rangle$. Then

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 2 & 0 \\ 2 & 4 & 1 \end{vmatrix} \\
&= \mathbf{i} \begin{vmatrix} 2 & 0 \\ 4 & 1 \end{vmatrix} - \mathbf{j} \begin{vmatrix} -1 & 0 \\ 2 & 1 \end{vmatrix} + \mathbf{k} \begin{vmatrix} -1 & 2 \\ 2 & 4 \end{vmatrix} \\
&= 2\mathbf{i} + \mathbf{j} - 8\mathbf{k} \\
&= \langle 2, 1, -8\rangle .
\end{aligned}$$

□

Remember that we said the cross product is our tool for finding perpendicular vectors. So it might be nice if we made sure it actually did that. As a quick check, we'll compute $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b})$ and $\mathbf{b} \cdot (\mathbf{a} \times \mathbf{b})$, with these vectors coming from (III.2). If we do so, we obtain

$$\begin{aligned}
\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) &= -2 + 2 = 0 \\
\mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) &= 4 + 4 - 8 = 0
\end{aligned}$$

Since these dot products are zero, this means that both $\mathbf{a}$ and $\mathbf{b}$ are perpendicular to the cross product $\mathbf{a} \times \mathbf{b}$. This is also true in general.

Theorem 11.4.4 Orthogonality of the Cross Product. $\mathbf{a} \times \mathbf{b}$ *is always orthogonal to both* $\mathbf{a}$ *and* $\mathbf{b}$*.*

So the cross product always produces orthogonal vectors. To determine the direction of the cross product $\mathbf{a} \times \mathbf{b}$, we use the **right-hand rule**: sweep your right hand from $\mathbf{a}$ to $\mathbf{b}$ and stick your thumb up. Then $\mathbf{a} \times \mathbf{b}$ is parallel to your thumb.

We would also like to know the magnitude of the cross product. We can just compute it using Definition 11.4.1 and find the magnitude using our usual formula. If we do so, we obtain (after a lot of simplifying!)

$$\|\mathbf{a} \times \mathbf{b}\|^2 = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \sin^2 \theta,$$

which reduces to the following result.

Theorem 11.4.5 Magnitude of the Cross Product. *Let θ denote the acute angle between the vectors* $\mathbf{a}$ *and* $\mathbf{b}$*, so that $0 \le \theta \le \pi$. Then*

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin \theta.$$

So in particular, two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ are parallel (i.e. have $\theta = 0$) if and only if $\mathbf{a} \times \mathbf{b} = \mathbf{0}$.

Example 11.4.6 Testing collinearity. We say that three points P, Q and R are **collinear** if they all lie on the same line. Suppose we want to check if the points $P(-2, 0, 1), Q(-1, 2, 4)$ and $R(1, 6, 10)$ are collinear or not. How can we do so? If we start by defining

$$\mathbf{u} = \vec{PQ} = \langle 1, 2, 3\rangle \quad \text{and} \quad \mathbf{v} = \vec{QR} = \langle 2, 4, 6\rangle ,$$

then we can say that all three points lie on the same line if and only if $\mathbf{u}$ and $\mathbf{v}$ are parallel to each other. So we'll compute their cross product to get $\mathbf{u} \times \mathbf{v} = \mathbf{0}$. Since these vectors are parallel, then the three given points must lie on the same line. □

Another important property of the magnitude of the cross product is the following: $\|\mathbf{a}\times\mathbf{b}\|$ is exactly equal to the area of the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$.

Example 11.4.7 Area of a triangle. Suppose that we want the area of the triangle with vertices $P(1,2,-1), Q(2,0,0)$ and $R(1,-1,-2)$. To start, we need to find vectors that determine the triangle. We can use

$$\mathbf{u} = \vec{PQ} = \langle 1,-2,1\rangle \quad \text{and} \quad \mathbf{v} = \vec{PR} = \langle 0,-3,-1\rangle\,.$$

Now, the triangle determined by $\mathbf{u}$ and $\mathbf{v}$ is precisely half of the parallelogram determined by these same vectors, so the area of this triangle is equal to $\frac{1}{2}\|\mathbf{u}\times\mathbf{v}\|$. We can use Sage as in the cell below to find the cross product of these vectors. Doing so, we get $\mathbf{u}\times\mathbf{v} = \langle 5,1,-3\rangle$. So the area of the triangle with vertices P, Q and R is

$$\begin{aligned}\text{area}\,\triangle PQR &= \frac{1}{2}\|\mathbf{u}\times\mathbf{v}\| \\ &= \frac{1}{2}\|\langle 5,1,-3\rangle\| \\ &= \frac{1}{2}\sqrt{35}.\end{aligned}$$

□

```
u = vector([1,-2,1])
v = vector([0,-3,-1])
u.cross_product(v)
```

```
(5, 1, -3)
```

Theorem 11.4.8 Properties of the Cross Product. *Let $\mathbf{a},\mathbf{b},\mathbf{c}$ be vectors and c a scalar. Then the following properties are true:*

1. $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$
2. $(c\mathbf{a})\times\mathbf{b} = c(\mathbf{a}\times\mathbf{b}) = \mathbf{a}\times(c\mathbf{b})$
3. $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$
4. $(\mathbf{a}+\mathbf{b})\times\mathbf{c} = \mathbf{a}\times\mathbf{c}+\mathbf{b}\times\mathbf{c}$
5. $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$

The last item above is an important relationship between the cross product and dot product called the **scalar triple product**. There is an important geometric significance to this new product.

Theorem 11.4.9 Geometry of the Scalar Triple Product. *Let $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$ be vectors in $\mathbb{R}^3$. Then $|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|$ is equal to the volume of the parallelepiped determined by $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$.*

Example 11.4.10 Testing if vectors are coplanar. We say that three vectors $\mathbf{u},\mathbf{v}$ and $\mathbf{w}$ are **coplanar** if they can all lie in a single plane. For example, $\mathbf{i},\mathbf{k}$ and $\langle -4,0,56\rangle$ are coplanar since they lie in the xz-plane, but $\mathbf{i},\mathbf{j}$ and $\mathbf{k}$ are not coplanar. Suppose we're given $\mathbf{u} = \langle -3,2,1\rangle, \mathbf{v} = \langle 1,2,0\rangle$ and $\mathbf{w} = \langle 1,1,1\rangle$. These vectors are coplanar if and only if the parallelepiped determined by these vectors has zero volume (i.e. is flat). Since

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = -9,$$

these vectors are *not* coplanar. □

Example 11.4.11 Another way to compute cross products. Using

Theorem 11.4.8 and the facts that

$$\begin{aligned}\mathbf{i}\times\mathbf{j}&=\mathbf{k}\\ \mathbf{j}\times\mathbf{k}&=\mathbf{i}\\ \mathbf{k}\times\mathbf{i}&=\mathbf{j}\\ \mathbf{a}\times\mathbf{a}&=\mathbf{0}\end{aligned}$$

gives us another way to compute cross products that doesn't involve determinants. As an example, let $\mathbf{u}=\langle 1,4,-3\rangle$ and $\mathbf{v}=\langle 0,2,1\rangle$. Then

$$\begin{aligned}\mathbf{u}\times\mathbf{v}&=(\mathbf{i}+4\mathbf{j}-3\mathbf{k})\times(2\mathbf{j}+\mathbf{k})\\ &=2\mathbf{i}\times\mathbf{j}+\mathbf{i}\times\mathbf{k}+8\mathbf{j}\times\mathbf{j}+4\mathbf{j}\times\mathbf{k}-6\mathbf{k}\times\mathbf{j}-3\mathbf{k}\times\mathbf{k}\\ &=2\mathbf{k}-\mathbf{j}+4\mathbf{i}-6\mathbf{i}\\ &=\langle -2,-1,2\rangle .\end{aligned}$$

□

Torque

Consider a force $\mathbf{F}$ acting on a rigid body at some position $\mathbf{r}$. This force applies a turning affect to the body, that we measure by **torque**.

Definition 11.4.12 Torque. The torque $\boldsymbol{\tau}$ of a force $\mathbf{F}$ acting at a position $\mathbf{r}$ is defined to be the vector

$$\boldsymbol{\tau}=\mathbf{r}\times\mathbf{F}.$$

♢

As one example of torque, consider a wrench applied to a bolt. The force is exerted at the end of the wrench, and the torque is a vector that's parallel to the axis of rotation of the bolt. The torque is greater if the force is applied at a direction perpendicular (or nearly so) to that of the wrench, and smaller if the force is nearly parallel to the direction of the wrench. As a quick check, the torque is $\mathbf{0}$ if the force is exactly parallel to the direction of the wrench, which makes sense: if we're pushing or pulling the wrench, the bolt won't rotate at all.

Example 11.4.13 Torque and hex keys. A hex key (Allen wrench) with a short arm of length 27 mm and a long arm length of 154 mm is applied to a screw, with the short arm attached to the screw. To turn the screw, a force of 0.5 N is applied to the long arm of the hex key turning the screw clockwise, and is exactly perpendicular to both the short arm and long arm of the hex key. We want to find the torque of this force on the screw.

One way we can do this is to imagine the screw sitting at the origin, and the hex key is (initially) in the xy-plane. Note that once the force is applied, it will begin to rotate the hex key out of the xy-plane. Now, the torque is defined by

$$\boldsymbol{\tau}=\mathbf{r}\times\mathbf{F}$$

where $\mathbf{r}$ is the vector from the screw to the point where the force $\mathbf{F}$ is applied. We can find $\mathbf{r}$ without too much trouble: it's $\mathbf{r}=\langle .027,.154,0\rangle$. To find the force $\mathbf{F}$, note that it's perpendicular to both the long arm and short arm of the screw. So a starting guess would be

$$\mathbf{F}=\langle .027,0,0\rangle\times\langle 0,.154,0\rangle=\langle 0,0,.042\rangle .$$

Such a force would turn the screw clockwise, but it has the wrong magnitude.

So we need to adjust it a bit: $\mathbf{F} = \langle 0, 0, .5\rangle$. So the torque is given by

$$\begin{aligned}\tau &= \mathbf{r} \times \mathbf{F} \\ &= \langle .027, .154, 0\rangle \times \langle 0, 0, .5\rangle \\ &= \langle .154, -.027, 0\rangle .\end{aligned}$$

□

SUGGESTED PROBLEMS: 1-19 odd, 29, 41

11.5 Equations of Lines and Planes

Recall that the equation of a line in $\mathbb{R}^2$ has the general form $ax + by + c = 0$. However, this equation does not give a line in $\mathbb{R}^3$. In fact, it actually gives a plane, as we'll see! What we want to do now is find exactly how to describe a given line or plane.

Equations of Lines

First, suppose we have a line L in $\mathbb{R}^3$ as follows:

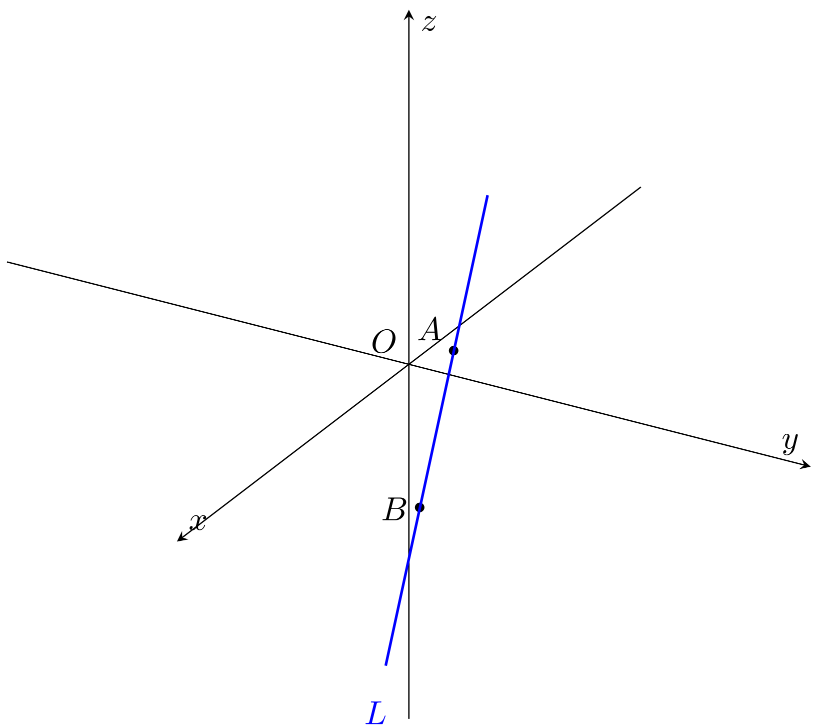


Figure 11.5.1 A line in $\mathbb{R}^3$.

Our goal is to find an equation that describes this line. Imagine that we have a vector $\mathbf{r}_0$ pointing from the origin to the point A, like so:

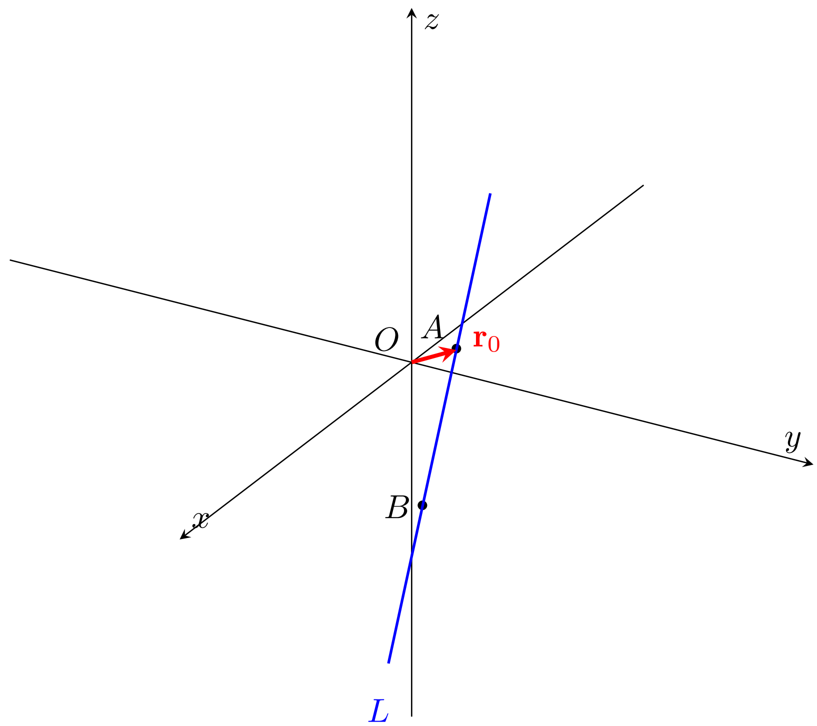


Figure 11.5.2 Tracing out L using vectors.

Then we can get to every other point on the line by starting from $\mathbf{r}_0$. In particular, if $\mathbf{u}$ is any vector parallel to $\vec{AB}$, then $\mathbf{r} = \mathbf{r}_0 + \mathbf{u}$ must also be on the line L:

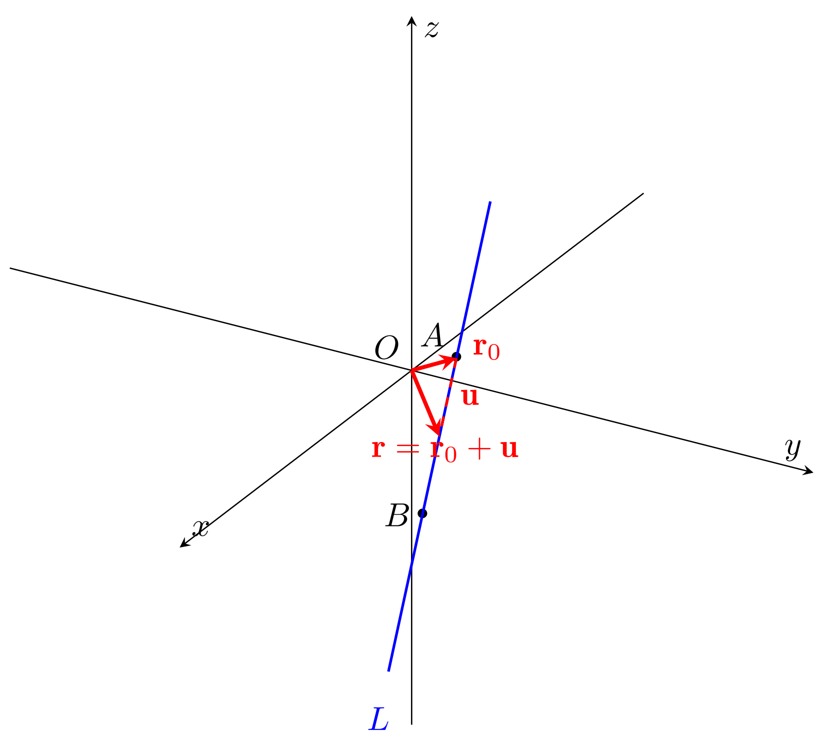


Figure 11.5.3 Tracing out L using vectors.

So to get all possible points on L, we just need to look at all possible vectors

$\mathbf{r} = \mathbf{r}_0 + \mathbf{u}$, where $\mathbf{u}$ is any vector parallel to $\vec{AB}$. If we set $\mathbf{v} = \vec{AB}$,[8] then

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$$

will trace out all points on L as t varies from $-\infty$ to ∞.

Theorem 11.5.4 Parametric Equations of a Line. *If $P_0(x_0, y_0, z_0)$ is a point on a line L, and if $\mathbf{v} = \langle a, b, c\rangle$ is parallel to L, then*

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$$

traces out L, where

$$\mathbf{r} = \langle x, y, z\rangle, \quad \mathbf{r}_0 = \langle x_0, y_0, z_0\rangle.$$

We can also write this as the system of equations

$$x = x_0 + ta, \quad y = y_0 + tb \quad \textit{and} \quad z = z_0 + tc.$$

Example 11.5.5 Equation of a line with a given direction and point. Suppose we want to find the equation of a line L that passes through $(2, 5, -3)$ and is perpendicular to $\mathbf{u} = \langle 2, 1, 2\rangle$. Theorem 11.5.4 says that we only need to find the right choices for $\mathbf{r}_0$ and $\mathbf{v}$ in order to get the equation of such a line, and we can write down one of these right away:

$$\mathbf{r}_0 = \langle 2, 5, -3\rangle.$$

So we just need to find $\mathbf{v}$. Since L is orthogonal to $\mathbf{u} = \langle 2, 1, 2\rangle$, this means we want $\mathbf{v}$ to be orthogonal to $\mathbf{u}$ as well. A great way to find such a $\mathbf{v}$ is to use the cross product. If $\mathbf{w}$ is any nonzero vector, then $\mathbf{u} \times \mathbf{w}$ must be orthogonal to both $\mathbf{u}$ and $\mathbf{w}$, and in particular it must be orthogonal to $\mathbf{u}$. So we can take

$$\mathbf{v} = \langle 2, 1, 2\rangle \times \mathbf{i} = \langle 0, 2, -1\rangle.$$

So the equation of this line is

$$\mathbf{r} = \langle 2, 1 + 2t, 2 - t\rangle$$

As t varies, this will trace out our line. □

When we begin working with line integrals, we'll need to know how to parameterize a line segment, so we'll discuss that now. Suppose we have a line segment between two points P_0 and P_1. Let $\mathbf{r}_0$ and $\mathbf{r}_1$ denote the corresponding position vectors of these points (i.e. the vectors whose components are given by the coordinates of these points). Then the equation of the *line* between P_0 and P_1 is given by

$$\mathbf{r}] = \mathbf{r}_0 + t(\mathbf{r}_1 - \mathbf{r}_0) = (1 - t)\mathbf{r}_0 + t\mathbf{r}_1.$$

To get the line segment we're actually looking for, we just need to restrict t to the interval $0 \leq t \leq 1$. So we can say that

$$\mathbf{r} = (1 - t)\mathbf{r}_0 + t\mathbf{r}_1 \quad \text{with} \quad 0 \leq t \leq 1$$

traces out the line segment between P_0 and P_1.

[8]We don't actually have to pick $\mathbf{v} = \vec{AB}$, we just need $\mathbf{v}$ to be parallel to $\vec{AB}$.

Equations of Planes

Now we move on to finding the equation of a plane. We'll try a vector approach to this like we did for lines, and this won't be too bad to do. We just need to keep the following in mind:

Observation 11.5.6 A plane in $\mathbb{R}^3$ is uniquely determined by specifying a single point for it to pass through and a direction for it to face.

So in particular, vectors will provide a nice way to specify the direction of a plane, and any vector that does so must be orthogonal to the plane. If a (nonzero) vector $\mathbf{n}$ is orthogonal to a given plane, then we call $\mathbf{n}$ a **normal vector** to the plane. Now we can build up an equation of a given plane.

Let $P_0(x_0, y_0, z_0)$ denote a point on the plane, and let $\mathbf{n} = \langle a, b, c\rangle$ be a normal vector to the plane. If $\mathbf{r} = \langle x, y, z\rangle$ is the position vector of any other point on the plane, then it must be true that $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$, where $\mathbf{r}_0$ is the position vector of P_0. We can summarize this in the following theorem.

Theorem 11.5.7 Vector and Scalar Equations of a Plane. *If $\mathbf{r}_0 = \langle x_0, y_0, z_0\rangle$ is the position vector for some point on a plane with normal vector $\mathbf{n} = \langle a, b, c\rangle$, then the position vector $\mathbf{r} = \langle x, y, z\rangle$ of any other point on the plane must satisfy the **vector equation of the plane**:*

$$\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0.$$

*Equivalently, any point (x, y, z) must also satisfy the equivalent **scalar equation of the plane**:*

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

Example 11.5.8 Equation of a plane orthogonal to a line. Consider the two lines in $\mathbb{R}^3$ determined by the equations

$$\begin{aligned}\mathbf{r}_1 &= \langle 1, 1, 0\rangle + t\,\langle 1, -1, 2\rangle\\ \mathbf{r}_2 &= \langle 2, 0, 2\rangle + s\,\langle -1, 1, 0\rangle\end{aligned}$$

We want to find the equation of the plane that contains the intersection of these two lines and is orthogonal to the first line. By Theorem 11.5.7, we need a point on the plane and a normal vector. Since the plane must contain the intersection of these two lines, that's our point. To find it, we'll set $\mathbf{r}_1 = \mathbf{r}_2$ and go from there:

$$\begin{aligned}\mathbf{r}_1 = \mathbf{r}_2 &\Rightarrow \langle -1, 1, -2\rangle + \langle t + s, s - t, 2t\rangle = \mathbf{0}\\ &\Rightarrow t + s - 1 = s - t + 1 = 2t - 2 = 0\\ &\Rightarrow t = 1 \quad \text{and} \quad s = 0\end{aligned}$$

So the position vector $\mathbf{r}_0$ of the point where these lines intersect is given by $\mathbf{r}_0 = \langle 2, 0, 2\rangle$. We also need a normal vector to the plane, but this is easy to find: since the plane needs to be orthogonal to the first line, we can just take $\mathbf{n} = \langle 1, -1, 2\rangle$. So the equation of our plane is

$$\langle 1, -1, 2\rangle \cdot \langle x - 2, y, z - 2\rangle = 0$$

or equivalently

$$x - y + 2z - 6 = 0.$$

This plane is plotted below using the Sage cell immediately after this example.

□

```
from sage.manifolds.utilities import set_axes_labels
var('x y')
P = plot3d((y-x)/2+3, (x,-3,3), (y,-3,3))
set_axes_labels(P,'x','y','z')
```

We can rewrite the equation of a plane in the form

$$ax + by + cz + d = 0.$$

A normal vector to this plane is given by $\mathbf{n} = \langle a, b, c\rangle$.

Example 11.5.9 Equation of a line orthogonal to a plane. Suppose we're given the plane $3x - 5z + 2 = 0$, and we want to find the equation of the line orthogonal to this plane and passing through the point $(3, -1, 2)$. Then the equation of this line is

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$$

where $\mathbf{r}_0 = \langle 3, -1, 2\rangle$ and $\mathbf{v}$ is a vector parallel to the line. Since the line must be orthogonal to the plane, we can take $\mathbf{v} = \langle 3, 0, -5\rangle$. So the equation of this line is

$$\mathbf{r} = \langle 3, -1, 2\rangle + \langle 3t, 0, -5t\rangle .$$

□

Example 11.5.10 Equation of a plane containing a line and a point. Suppose we're trying to find a plane containing the point $(1, 2, 0)$ and containing the line $\mathbf{r} = \langle 3 - t, 2, 5 + 2t\rangle$. To get the equation of this plane, we also need a normal vector to the plane. We can find the normal vector by computing the cross product of two vectors parallel to the plane.

First, since the plane contains the line $\langle 3 - t, 2, 5 + 2t\rangle$, the direction of this line must be parallel to the plane. So we know that $\mathbf{v} = \langle -1, 0, 2\rangle$ is parallel to the plane. To get another vector parallel to this plane, we can use the given point and a second point in the plane, say $(3, 2, 5)$, which comes from the line (just set $t = 0$). So the vector $\mathbf{u} = \langle 3 - 1, 2 - 2, 5 - 0\rangle = \langle 2, 0, 5\rangle$ must also be parallel to the plane. Hence a normal vector to this plane is

$$\mathbf{n} = \mathbf{u} \times \mathbf{v} = \langle 0, 9, 0\rangle .$$

So an equation of this plane is given by

$$9(y - 2) = 0.$$

□

11.6 Cylinders and Quadric Surfaces

Before we move on to doing calculus with vectors, we'll briefly take a look at more graphs in $\mathbb{R}^3$. In particular, we'll look at **cylinders** and **quadric surfaces**.

Cylinders

Definition 11.6.1 Cylinders. A cylinder is the collection of all lines parallel to a given line and passing through some plane curve. ♢

A basic example of a cylinder is the set of all lines passing through the unit circle in the xy-plane and parallel to the z-axis. In $\mathbb{R}^3$, this is just the graph of $x^2 + y^2 = 1$. The graph of this cylinder is provided below.

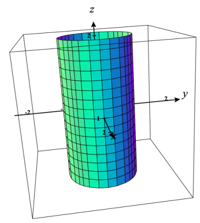


Figure 11.6.2 The cylinder $x^2 + y^2 = 1$

For our purposes, equations that give a cylinder will often be missing a variable.

Example 11.6.3 A sinusoidal cylinder. Consider the equation $z = \sin y$ in $\mathbb{R}^3$. This equation is missing the variable x, which suggests that the graph of this equation should be a cylinder. However, it's not going to look like the cylinders we may be used to at this point. In fact, this is just the set of all lines passing through the curve $z = \sin y$ in the yz-plane and parallel to the x-axis. It's graph is given below. □

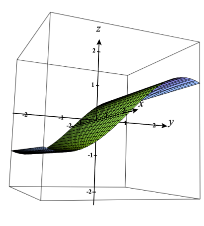


Figure 11.6.4 The cylinder $z = \sin(y)$

Example 11.6.5 Another cylinder. Consider the cylinder given by the set of all lines passing through the plane curve $y = 3x - 1$ in the xy-plane and parallel to the line in $\mathbb{R}^3$ defined by the equation

$$\mathbf{r} = \langle 3 - t, t, 3 \rangle .$$

What does this cylinder look like? Well, we can view it as essentially a "sheet" of lines cutting through the xy-plane at the line $y = 3x - 1$. If we try to imagine this, then this suggests that this cylinder should probably be a plane! In fact, this cylinder is exactly the plane containing the point $(0, -1, 0)$ and parallel to the line $\mathbf{r}$ given above. As the line itself is parallel to the xy-plane, the resulting cylinder is just the xy-plane. □

Quadric Surfaces

A **quadric surface** is any surface that is the graph of an equation of the form

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + 0 = 0.$$

A useful tool for graphing quadric surfaces (and others in $\mathbb{R}^3$) is the concept of a **trace**, which is what the curve looks like in a plane parallel to the one the coordinate planes. This amounts to setting either, x, y or z equal to a constant and graphing the resulting equation.

Example 11.6.6 An ellipsoid. Consider the equation

$$4x^2 + \frac{y^2}{9} + \frac{z^2}{25} = 1.$$

If we want to graph this, we can graph a few of it's traces to get an idea of what it looks like. Let's graph traces parallel the xz-plane to start. This means we'll set y equal to different constants. For $y = 0$, we get the equation

$$4x^2 + \frac{z^2}{25} = 1$$

which we rewrite as

$$\frac{x^2}{(\frac{1}{2})^2} + \frac{z^2}{25} = 1.$$

This is an ellipse in the xz-plane, with minor axis $\frac{1}{2}$ and major axis 5. We can graph another trace, say in the xy-plane, we get

$$\frac{x^2}{(\frac{1}{2})^2} + \frac{y^2}{9} = 1,$$

which is an ellipse with minor axis $\frac{1}{2}$ and major axis 3. Similarly, in the yz-plane we have an ellipse with minor axis 3 and major axis 5. Putting these together gives us a rough idea of the shape of this surface, which we call an **ellipsoid**. □

Example 11.6.7 Region between surfaces. Suppose we want to sketch the region between the surface $z = \sqrt{x^2 + y^2}$ and the cylinder $x^2 + y^2 = 1$ for $0 \leq z \leq 1$. First, we can graph $z = \sqrt{x^2 + y^2}$. If we look at the horizontal traces of this surface, we get circles of varying radii. As z increases, the radii of these circles increase as well. This surface is just a cone! So we're describing the region of this cone bounded between $z = 0$ and $z = 1$, and contained inside the cylinder $x^2 + y^2 = 1$. □

Figure 11.6.8 The region contained between $z = \sqrt{x^2 + y^2}$ and $x^2 + y^2 = 1$

11.7 Vector Functions

Recall from Equations of Lines that the equation of a line can be written as

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v} = \langle x_0 + at, y_0 + bt, z_0 + ct \rangle .$$

This is our first example of a **vector function**. Vector functions are functions of the form

$$\mathbf{r}(t) = \langle f(t), g(t), h(t) \rangle ,$$

and graphs of vector functions are called **space curves**. We call f, g, h the **component functions** of $\mathbf{r}$. We're interested in how these curves change, which means we're interested in how to do calculus on space curves. Although these curves live in $\mathbb{R}^3$, there's still only one independent variable: t. So much of what we learned in Calculus I applies to space curves.

Limits with Space Curves

We can take limits with vector functions just as we can with regular functions.

Let $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle$. Then

$$\lim_{t\to a} \mathbf{r}(t) = \left\langle \lim_{t\to a} f(t), \lim_{t\to a} g(t), \lim_{t\to a} h(t) \right\rangle. \tag{III.3}$$

In other words, if you want to take the limit of a vector function you can just take the limits of the component functions.

Example 11.7.1 Limit of a vector function. Let

$$\mathbf{r}(t) = \frac{t^2-1}{t-1}\mathbf{i} + \sqrt{t+8}\mathbf{j} - \frac{\sin \pi t}{\ln t}\mathbf{k}.$$

Suppose we want to find $\lim_{t\to 1} \mathbf{r}(t)$. Then we just need to take the limit of each component. So

$$\lim_{t\to 1} \mathbf{r}(t) = 2\mathbf{i} + 3\mathbf{j} - \pi\mathbf{k}.$$

□

Just as in Calculus I, we say that a vector function $\mathbf{r}(t)$ is **continuous** at $t = a$ if $\lim_{t\to a} \mathbf{r}(t) = \mathbf{r}(a)$. In general, a vector function is continuous wherever *all* of its components functions are continuous.

Example 11.7.2 A horiztonal helix. Let $\mathbf{r}(t) = \langle \sin \pi t, t, \cos \pi t\rangle$, and suppose we want to sketch this function. One way to do so is to plug in values for t and connect the resulting points with a curve, but we can also do the following to get an idea of what this looks like. First, note that we have $x = \sin \pi t$ and $z = \cos \pi t$. So $x^2 + z^2 = 1$, which means that this looks like the unit circle in the xz-plane, traced *clockwise*. Since we also have $y = t$, this curve moves farther along the y-axis as t increases. If we trace this out, we get a helix (see the below plot). We can also see from the graph that it has no jumps or gaps, so $\mathbf{r}(t)$ is continuous everywhere. □

```
# Code shamelessly adapted from this stackoverflow post:
    http://stackoverflow.com/questions/26989131/add-cylinder-to-plot
import matplotlib.pyplot as plt
import numpy as np
from mpl_toolkits.mplot3d import Axes3D

plt.rc('text', usetex=True)

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

# Define variables/components
t = np.linspace(-5,5,500)
x = np.sin(np.pi*t)
y = t
z = np.cos(np.pi*t)

# Draw helix
ax.plot(x, y, z, label=r'$\mathbf{r}(t)$')
ax.legend()

ax.set_xlabel(r'$x$')
ax.set_ylabel(r'$y$')
ax.set_zlabel(r'$z$')
plt.show()
```

Example 11.7.3 Finding vector functions. Consider the cylinder $x^2 + y^2 = 4$ and the surface $z = xy$, and suppose we want to trace out there intersection with a vector function. Here's how we can do this. First, we'll come up with the x and y components of $\mathbf{r}(t)$. Since $x^2 + y^2 = 4$, this suggests that we should take

$$x = 2\cos t$$
$$y = 2\sin t$$

So that's two down, one to go. To get z, we just need to use the equation $z = xy$. So

$$z = xy = 4\cos t \sin t.$$

So our vector function is

$$\mathbf{r}(t) = \langle 2\cos t, 2\sin t, 2\sin 2t\rangle .$$

This is also plotted below. □

```
# Code shamelessly adapted from this stackoverflow post:
    http://stackoverflow.com/questions/26989131/add-cylinder-to-plot
# If I change it enough it becomes MY code
import matplotlib.pyplot as plt
import numpy as np
from mpl_toolkits.mplot3d import Axes3D

plt.rc('text', usetex=True)

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

# Define variables/components
t = np.linspace(-5,5,500)
x = 2*np.cos(t)
y = 2*np.sin(t)
z = x*y

# Draw curve
ax.plot(x, y, z, label=r'$\mathbf{r}(t)$')
ax.legend()

ax.set_xlabel(r'$x$')
ax.set_ylabel(r'$y$')
ax.set_zlabel(r'$z$')
plt.show()
```

Derivatives with Space Curves

Now that we know how to take limits with vector functions, we can take derivatives as well.

Definition 11.7.4 Derivatives of Vector Functions. Let $\mathbf{r}(t)$ denote a vector function. The **derivative** of $\mathbf{r}$ is the new vector function $\mathbf{r}'$ given by

$$\mathbf{r}'(t) = \frac{d\mathbf{r}}{dt} = \lim_{h\to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h},$$

assuming that the limit exists. If this limit exists, we say that $\mathbf{r}$ is **differentiable**. ♢

If $\mathbf{r}(t) = \langle f(t), g(t), h(t) \rangle$, then $\mathbf{r}$ is differentiable if and only if f, g, h are, and $\mathbf{r}'(t) = \langle f'(t), g'(t), h'(t) \rangle$. Just as in Calculus I, the derivative represents how quickly a space curve is changing at some value of t. However, derivatives of vector functions also carry information about the *direction* a curve is moving. We call $\mathbf{r}'(t)$ the **tangent vector** to $\mathbf{r}(t)$. In particular, $\mathbf{r}'(t)$ is parallel to the space curve $\mathbf{r}$ at t, and its magnitude $\|\mathbf{r}'(t)\|$ represents how quickly the curve is changing at t. If we only care about direction, then we can define the **unit tangent** $\mathbf{T}(t)$, which is given by

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}.$$

We also have the usual ideas from Calculus I and physics regarding motion: velocity is the derivative of position and acceleration is the derivative of velocity.

Example 11.7.5 Velocity on a saddle. A particle moves counterclockwise along the "saddle" $\mathbf{r}(t) = \langle 2\cos t, 2\sin t, 2\sin 2t \rangle$. We want its velocity at $t = \frac{\pi}{2}$. First, find $\mathbf{r}'$ to get

$$\mathbf{r}'(t) = \langle -2\sin t, 2\cos t, 4\cos 2t \rangle .$$

At $t = \frac{\pi}{2}$, we have the velocity vector

$$\mathbf{r}'\left(\frac{\pi}{2}\right) = \langle -2, 0, -4 \rangle .$$

So at the point $t = \frac{\pi}{2}$, the space curve is parallel to the vector $\langle -2, 0, 4 \rangle$. In other words, the particle is moving in this direction at $t = \frac{\pi}{2}$. □

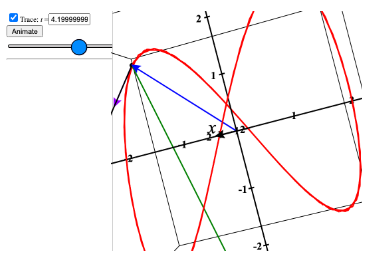


Figure 11.7.6 Motion along the saddle traced by $\mathbf{r}(t)$ in Example 11.7.5

Example 11.7.7 Tangents on a circle. A particle moves along the circle $y^2 + z^2 = 5$ in the yz-plane, counterclockwise and with an angular frequency of $5\pi\,\frac{\text{rad}}{\text{s}}$. Then we can assume that its position is described by

$$\mathbf{r}(t) = \left\langle 0, \sqrt{5}\cos 5\pi t, \sqrt{5}\sin 5\pi t \right\rangle .$$

Suppose we want to find the direction this particle is going at any given moment. Then we can just find the unit tangent vector $\mathbf{T}$:

$$\mathbf{T}(t) = \frac{\langle 0, -5\sqrt{5}\sin 5\pi t, 5\sqrt{5}\cos 5\pi t \rangle}{\sqrt{125}}.$$

□

We also have derivative rules for vector functions, based off of the familiar formulas from Calculus I.

Theorem 11.7.8 Vector Derivative Rules. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be differentiable vector functions,* c *be a scalar and let* $f(t)$ *be a differentiable (scalar) function.*

Then the following formulas hold:

1. $\dfrac{d}{dt}[\mathbf{u}(t)+\mathbf{v}(t)] = \mathbf{u}'(t)+\mathbf{v}'(t)$

2. $\dfrac{d}{dt}[c\mathbf{u}(t)] = c\mathbf{u}'(t)$

3. $\dfrac{d}{dt}[f(t)\mathbf{u}(t)] = f'(t)\mathbf{u}(t)+f(t)\mathbf{u}'(t)$

4. $\dfrac{d}{dt}[\mathbf{u}(t)\cdot\mathbf{v}(t)] = \mathbf{u}'(t)\cdot\mathbf{v}(t)+\mathbf{u}(t)\cdot\mathbf{v}'(t)$

5. $\dfrac{d}{dt}[\mathbf{u}(t)\times\mathbf{v}(t)] = \mathbf{u}'(t)\times\mathbf{v}(t)+\mathbf{u}(t)\times\mathbf{v}'(t)$

6. $\dfrac{d}{dt}\mathbf{u}(f(t)) = f'(t)\mathbf{u}'(f(t))$

Integrals with Space Curves

We can also integrate vector functions without too much trouble. Just as taking the derivative of a vector function reduces down to differentiating each component, integrating a vector function reduces down to integrating each component. If $\mathbf{r}(t) = \langle f(t), g(t), h(t)\rangle$, then

$$\int_a^b \mathbf{r}(t)\,dt = \left\langle \int_a^b f(t)\,dt, \int_a^b g(t)\,dt, \int_a^b h(t)\,dt \right\rangle. \tag{III.4}$$

SUGGESTED PROBLEMS: 1, 3, 5, 7, 17, 21, 23, 37, 41, 45, 59, 61

11.8 Arc Length and Curvature

Arc Length

Suppose some particle travels along the space curve given by $\mathbf{r}(t)$ for $a \le t \le b$. What we'd like to do is determine how far the particle travels over this time interval. Recall that distance is just speed multiplied by time, and the speed of this particle is given by $\|\mathbf{r}'(t)\|$. So we can imagine that the distance traveled by this particle over some infinitesimal time interval dt to be given by $\|\mathbf{r}'(t)\|\ dt$. So adding up all of these distances from $t = a$ to $t = b$ should give us the arc length. This suggests (but does not prove!) the following formula for arc length:

$$L = \int_a^b \|\mathbf{r}'(t)\|\ dt. \tag{III.5}$$

(III.5) gives the length L of the space curve $\mathbf{r}(t)$ from $t = a$ to $t = b$, assuming that the integral exists.

Example 11.8.1 Arc length of a helix. Suppose we want to find the arc length of the helix $\mathbf{r}(t) = \langle \cos 3t, \sin 3t, t\rangle$ from $t = 0$ to $t = 1$. Then this is given by

$$L = \int_0^1 \|\langle -3\sin 3t, 3\cos 3t, 1\rangle\|\ dt = \int_0^1 \sqrt{10}\,dt = \sqrt{10}.$$

□

One thing we'd like to do now is to parametrize a space curve $\mathbf{r}(t)$ with respect to arc length. Here's what we mean by this: given some $\mathbf{r}(t)$ with $a \leq t \leq b$, we define its **arc length function** $s(t)$ by

$$s(t) = \int_a^t \|\mathbf{r}'(u)\| \, du.$$

If we can then solve for t in terms of s, this parametrizes the curve $\mathbf{r}(t)$ in terms of the arc length variable s.

Example 11.8.2 Reparametrizing a space curve. Suppose we're given the space curve

$$\mathbf{r}(t) = \left(\frac{2}{t^2+1} - 1\right)\mathbf{i} + \frac{2t}{t^2+1}\mathbf{j}$$

which starts at $(1, 0)$, (so t starts at 0) and we want to find the point that is $\frac{\pi}{2}$ units along the curve in the positive direction. Then we can do this by reparametrizing the curve using arc length. Here's how. First, we find the arc length function $s(t)$:

$$\begin{aligned}
s(t) &= \int_0^t \|\mathbf{r}'(u)\| \, du \\
&= \int_0^t \sqrt{\frac{4u^2 + 4(u^2+1)^2 - 16u^2(u^2+1) + 16u^4}{(u^2+1)^4}} \, du \\
&= \int_0^t \frac{2}{u^2+1} \, du \\
&= 2\tan^{-1} t
\end{aligned}$$

Since $s = 2\tan^{-1} t$, we get $t = t(s) = \tan \frac{s}{2}$. So

$$\mathbf{r}(t(s)) = \left(\frac{2}{\tan^2 \frac{s}{2} + 1} - 1\right)\mathbf{i} + \frac{2\tan \frac{s}{2}}{\tan^2 \frac{s}{2} + 1}\mathbf{j}$$

reparametrizes the space curve $\mathbf{r}$ in terms of arc length. So the point on the curve that is $\frac{\pi}{2}$ units along in the positive direction is given by

$$\mathbf{r}(t(\pi/2)) = \langle 0, 1 \rangle.$$

□

```
# Code shamelessly adapted from this stackoverflow post:
    http://stackoverflow.com/questions/26989131/add-cylinder-to-plot
# If I change it enough it becomes MY code
import matplotlib.pyplot as plt
import numpy as np
from mpl_toolkits.mplot3d import Axes3D

plt.rc('text', usetex=True)

fig = plt.figure()
ax = fig.add_subplot(111)

# Define variables/components
t = np.linspace(-5,5,500)
x = 2/(t^2+1) - 1
y = 2*t/(t^2+1)
```

```
# Draw curve
ax.plot(x, y, label=r'$\mathbf{r}(t)$')
ax.legend()

ax.set_xlabel(r'$x$')
ax.set_ylabel(r'$y$')
plt.show()
```

Curvature

What we want to do now is measure how much a curve ”turns” at some point. First, we call a space curve $\mathbf{r}(t)$ **smooth** on some interval I if $\mathbf{r}'(t) \neq 0$ for any t in I. Now, if we want to measure how quickly a curve is changing direction then we can use the unit tangent vector $\mathbf{T}(t)$ to measure this. In particular, the derivative of the unit tangent should provide a good measure of how quickly a curve turns. However, we don't want to let the specific parametrization of the curve affect this; in other words, the rate at which a curve is turning should not depend on the speed at which the curve is traveled. This leads to the following definition.

Definition 11.8.3 Curvature. Let $\mathbf{r}(t)$ denote a smooth curve on some interval I. The **curvature** of $\mathbf{r}(t)$ on I is defined to be the function $\kappa(t)$ given by

$$\kappa(t) = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{r}'(t)\|},$$

where $\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}$ is the unit tangent to the curve. ◊

Example 11.8.4 Computing a curvature. Consider the curve $\mathbf{r}(t) = \left\langle t, \frac{1}{2}t^2, t^2 \right\rangle$. We'll try to find the curvature by making use of Definition 11.8.3. First, we need to find $\mathbf{T}$ so we can take its derivative:

$$\mathbf{T}(t) = \frac{\langle 1, t, 2t \rangle}{\sqrt{1 + t^2 + 4t^2}}.$$

Therefore

$$\begin{aligned}
\mathbf{T}'(t) &= -\frac{1}{2}(1+5t^2)^{-3/2}(10t)\langle 1, t, 2t \rangle + (1+5t^2)^{-1/2}\langle 0, 1, 2 \rangle \\
&= -5t(1+5t^2)^{-3/2}\langle 1, t, 2t \rangle + (1+5t^2)^{-1/2}\langle 0, 1, 2 \rangle . \\
&= (1+5t^2)^{-3/2}\langle -5t, 1, 2 \rangle
\end{aligned}$$

So the curvature is given by

$$\begin{aligned}
\kappa(t) &= \frac{(1+5t^2)^{-3/2}\sqrt{25t^2+5}}{(1+5t^2)^{1/2}} \\
&= \sqrt{5}(1+5t^2)^{-3/2}
\end{aligned}$$

□

Example 11.8.5 Using curvature. Consider the curve $\mathbf{r}(t) = \left\langle t, \frac{1}{2}t^2, t^2 \right\rangle$ once again. We'll try to find where this curve turns the fastest. To do this, we look at the curvature $\kappa(t)$ we found from Example 11.8.4:

$$\kappa(t) = \sqrt{\frac{5}{(1+5t^2)^3}}.$$

This curve is turning the fastest precisely where the curvature is largest, and this happens at $t = 0$. □

We can use Definition 11.8.3 to compute curvatures, but we've seen that it can be pretty awful. So we'd like to use another formula if we can; thankfully, there is another option.

Theorem 11.8.6 Alternative Curvature Formula. *The curvature of* $\mathbf{r}(t)$ *in* $\mathbb{R}^3$ *is given by*

$$\kappa(t) = \frac{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|^3}.$$

The formula in Theorem 11.8.6 may look worse than the formula given in Definition 11.8.3, but it has one significant advantage: we don't need to differentiate any magnitudes, which is required in the previous formula to compute $\mathbf{T}'(t)$.

Example 11.8.7 Using the alternative formula. Let $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + e^t\mathbf{k}$. We'll make use of Theorem 11.8.6 to find $\kappa(t)$. We have

$$\begin{aligned}\mathbf{r}'(t) &= \left\langle 1, 2t, e^t \right\rangle \\ \mathbf{r}''(t) &= \left\langle 0, 2, e^t \right\rangle\end{aligned}$$

and so

$$\kappa(t) = \frac{\|\langle 2e^t(t-1), -e^t, 2\rangle\|}{\|\langle 1, 2t, e^t\rangle\|^3} = \frac{\sqrt{4e^{2t}(t-1)^2 + e^{2t} + 4}}{(1 + 4t^2 + e^{2t})^{3/2}}.$$

□

Normal and Binormal Vectors

Consider a curve $\mathbf{r}(t)$. Then the direction of the curve is given by the unit tangent vector $\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}$. If we differentiate $\mathbf{T}(t)$ to get $\mathbf{T}'(t)$, we can view this vector as telling us the direction the curve is turning. This leads us to the concept of **unit normal vectors** to a space curve.

Definition 11.8.8 Unit Normal Vectors. Consider a space curve given by $\mathbf{r}(t)$. The unit normal vector is the vector $\mathbf{N}(t)$ given by

$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|},$$

where $\mathbf{T}'$ is the derivative of the unit tangent vector. ◊

Example 11.8.9 Unit normal on a circle. Find the unit normal vector of the curve given by $\mathbf{r}(t) = \langle \cos t, \sin t\rangle$.

Solution. If we think of a particle moving along $\mathbf{r}(t)$, then this particle is just moving along the unit circle. So at every point along this path, the particle should be turning toward the origin in order to stay on the unit circle. So at all points of the curve, $\mathbf{N}(t)$ should point towards the origin. To prove this, we'll use the formula above to find the unit normal:

$$\begin{aligned}\mathbf{N}(t) &= \frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|} \\ &= \frac{\langle -\cos t, -\sin t\rangle}{1}\end{aligned}$$

So $\mathbf{N}(t) = \langle -\cos t, -\sin t\rangle = -\mathbf{r}(t)$. So at every point of the circle, the unit normal points in the opposite direction of the corresponding position vector, i.e. it points towards the origin. □

It's not clear from Definition 11.8.8, but $\mathbf{N}(t)$ is *always* orthogonal to $\mathbf{T}(t)$ for any space curve/vector function $\mathbf{r}(t)$. We can use these two vectors to get another vector called the **binormal vector**.

Definition 11.8.10 Binormal Vector. Let $\mathbf{r}(t)$ be a vector function with unit tangent $\mathbf{T}(t)$ and unit normal $\mathbf{N}(t)$. Then the unit binormal vector is the vector $\mathbf{B}(t)$ given by

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

◊

11.9 Motion in Space

We now move on to examining motion in $\mathbb{R}^3$. First, recall that $\mathbf{r}(t)$ represents the position vector of some particle moving along a curve in space, then its velocity $\mathbf{t}(t)$ and acceleration $\mathbf{a}(t)$ are given by

$$\mathbf{v}(t) = \mathbf{r}'(t)$$
$$\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t)$$

The speed of the particle is just the magnitude of the velocity: $\|\mathbf{v}(t)\|$.

Example 11.9.1 Motion of a projectile. A projectile is fired out of a cannon with an initial speed of $200\,\frac{\mathrm{m}}{\mathrm{s}}$ to the west and with an angle of elevation of 30°. If the particle was fired from a raised platform that is 50 m off level ground, where does the particle land?

Solution. First, we'll assume that $\mathbf{j}$ points northward and $\mathbf{k}$ points straight up. Let's assume that the platform is directly above the origin. If we let $\mathbf{r}(t)$ denote the position (in meters) of the particle at time t (in seconds), then we can say that $\mathbf{r}_0 = \mathbf{r}(0) = \langle 0, 0, 50\rangle$. We also have

$$\mathbf{v}(0) = -100\sqrt{3}\mathbf{i} + 100\mathbf{k}$$
$$\mathbf{a} = -9.8\mathbf{k}$$

We can integrate up to find the position $\mathbf{r}(t)$:

$$\mathbf{v}(t) = -9.8t\mathbf{k} + \mathbf{C}$$

where $\mathbf{C}$ is an arbitrary constant vector. To find it, we'll use our initial condition on $\mathbf{v}$:

$$\mathbf{v}(0) = \mathbf{C} = -100\sqrt{3}\mathbf{i} + 100\mathbf{k}.$$

So $\mathbf{v}(t) = -9.8t\mathbf{k} - 100\sqrt{3}\mathbf{i} + 100\mathbf{k}$. Integrating once more to get the position, we have

$$\mathbf{r}(t) = -100\sqrt{3}t\mathbf{i} + (100t - 4.9t^2)\mathbf{k} + \mathbf{D} = -100\sqrt{3}t\mathbf{i} + (100t - 4.9t^2)\mathbf{k} + 50\mathbf{k}.$$

So $\mathbf{r}(t) = \left\langle -100\sqrt{3}t, 0, 50 + 100t - 4.9t^2 \right\rangle$.

To find where the particle lands, we just set the third component equal to zero and solve for t to get

$$t = \frac{-100 \pm \sqrt{10000 + 980}}{-9.8} = -.4883, 20.896.$$

We need to choose the positive value for t, and if we do so we see that when the projectile hits ground it's at position $\mathbf{r}(20.896) = \langle -3619.38, 0, 0\rangle$. So the projectile is a little over 3.5 km to the west. □

To describe acceleration on a curve $\mathbf{r}(t)$, it can be useful to break it down into **tangential components** and **normal components**.

Definition 11.9.2 Components of Acceleration. Let $\mathbf{r}(t)$ denote a space curve with unit tangent $\mathbf{T}(t)$, unit normal $\mathbf{N}(t)$ and acceleration $\mathbf{a}$. The tangential component of acceleration is the scalar a_T given by

$$a_T = \mathbf{a} \cdot \mathbf{T}.$$

The normal component is the scalar a_N given by

$$a_N = \mathbf{a} \cdot \mathbf{N}.$$

◊

a_T represents how much of the acceleration is directed tangent to the curve $\mathbf{r}(t)$, whereas a_N represents how much of the acceleration is directed perpendicular to the tangent.

Example 11.9.3 Components of acceleration for a projectile. Consider the particle from Example 11.9.1. At what point is the tangential component of acceleration greatest?

Solution. Since $\mathbf{a} = -9.8\mathbf{k}$, the tangential component should be greatest when the projectile is fired or when it hits the ground, since these are the points where the direction of the trajectory most closely matches the direction of acceleration. To actually verify this, we'll find the tangential component using Definition 11.9.2. Since

$$\mathbf{T}(t) = \left\langle -\frac{100\sqrt{3}}{\sqrt{(100-9.8t)^2+30000}}, 0, \frac{100-9.8t}{\sqrt{(100-9.8t)^2+30000}} \right\rangle$$

it follows that

$$\begin{aligned} a_T &= \mathbf{a} \cdot \mathbf{T} \\ &= -9.8\frac{100-9.8t}{\sqrt{(100-9.8t)^2+30000}} \end{aligned}$$

So

$$\begin{aligned} |a_T| &= 9.8\frac{|100-9.8t|}{\sqrt{(100-9.8t)^2+30000}} \\ &= \frac{9.8}{\sqrt{1+\frac{30000}{(100-9.8t)^2}}} \end{aligned}$$

We can make this as large as possible by making $(100-9.8t)^2$ as large as possible, and this reaches its largest value at $t = 20.896$, when the projectile hits the ground. □

```
t = var('t')
r = vector([-100*sqrt(3)*t, 0, 50 + 100*t - 4.9*t^2])
T = derivative(r,t).normalized()
show(T)
```

Chapter 12

Partial Derivatives

For functions that depend on a single independent variable such as x or t, derivatives are useful because they represent how the function itself changes as the independent variable changes. However, there are many situations where it makes sense to consider functions of more than one independent variable. Therefore, we will want to extend the notion of derivative to include these functions as well. Our primary tool for this will be the **partial derivative**, which we'll see in Section 12.3. But we'll need to build up some terminology to get there.

12.1 Functions of Several Independent Variables

Now we need to consider calculus with functions of more than one independent variable.

Functions of Two Variables

A function of several variables is a function that depends on more than one independent variable. One such example comes from differential equations, where quantities such as the temperature of a metal rod depend on both a position x and a time t.

Definition 12.1.1 Functions of Two Variables. A **function of two variables** is a rule f or $z = f(x, y)$ that assigns pairs (x, y) of real numbers to real numbers. The **domain** is the set D in $\mathbb{R}^2$ for which the rule is defined. The **range** is the set R in $\mathbb{R}$ of all possible values of f. x, y are called independent variables and z is called the **dependent variable**. $\Diamond$

Example 12.1.2 Domain of a function. Let $f(x, y) = \frac{1}{\sqrt{4-x^2-y^2}}$. Find the domain of f.

Solution. We'll break this down into pieces. First, since we're dealing with square roots in real numbers we need $x^2 + y^2 \leq 4$. Second, we need to make sure that $\sqrt{4 - x^2 - y^2} \neq 0$. So the domain of f is the set of all points (x, y) in $\mathbb{R}^2$ such that $x^2 + y^2 < 4$. In other words, the domain is just the inside of the circle of radius 2 centered at the origin. $\square$

We can also graph functions of two variables by viewing $f(x, y)$ as representing the height of some surface over the point (x, y) in the xy-plane. In other words, just as $y = f(x)$ represents a curve in $\mathbb{R}^2$, the equation $z = f(x, y)$ represents a surface in $\mathbb{R}^3$.

Example 12.1.3 Sketching a cone. Sketch the surface given by $f(x,y) = \sqrt{4x^2+y^2}$.

Solution. We need to sketch $z = \sqrt{4x^2+y^2}$. We can start by squaring both sides to get $z^2 = 4x^2 + y^2$. If we look at the right hand side, it looks a lot like an ellipse, so we'll try to work with that. Assuming $z \neq 0$, we can rewrite the equation to get

$$\frac{x^2}{z^2/4} + \frac{y^2}{z^2} = 1$$

For any given positive value of z, this represents an ellipse that extends to $\frac{z}{2}$ in the x direction and z in the y direction. So as z increases, these ellipses increase in size as well. □

Contour Plots

Definition 12.1.4 Contour Plots. Given a function $f(x,y)$, a **contour plot** is the plot of an equation of the form $f(x,y) = k$ for some constant k. ♢

A contour plot can be viewed as slicing a surface in $\mathbb{R}^3$ with a level plane. The contour plot is exactly where the plane intersects the surface.

Example 12.1.5 Contours on a paraboloid. What do the contours of $f(x,y) = x^2 + \frac{y^2}{9}$ look like?

Solution. If we pick different (nonnegative) values for k, then all of the contours of $f(x,y)$ take the form

$$x^2 + \frac{y^2}{9} = k$$

which turns into

$$\frac{x^2}{k} + \frac{y^2}{9k} = 1.$$

This is a family of ellipses. □

SUGGESTED PROBLEMS: 1, 5, 11, 13

12.2 Limits and Continuity

In this section we begin extending calculus concepts from one dimension to higher dimensions, starting with limits.

Limits

Just as functions of one variable have a notion of limit, the same can be said for functions of several variables. However, limits involving several variables tend to be more restrictive.

Definition 12.2.1 Limit of a Function of Two Variables. Let $f(x,y)$ be some function and let (a,b) be a point arbitrarily close to the domain of f. We say that

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

if $f(x,y)$ gets arbitrarily close to L as (x,y) gets arbitrarily close to (a,b). ♢

Here's why limits can be tricky in two dimensions: for the limit to exist, $f(x,y)$ must approach the same value L no matter how (x,y) approaches (a,b)!

Example 12.2.2 Find

$$\lim_{(x,y)\to(0,0)} \frac{x^2 \sin y}{x^3 + 2y^3}$$

or show that it doesn't exist.

Solution. First, we'll start simple. If we set $x = 0$ then the limit becomes

$$\lim_{(x,y)\to(0,0)} \frac{x^2 \sin y}{x^3 + 2y^3} = \lim_{y\to 0} \frac{0}{2y} = 0.$$

Likewise, if we set $y = 0$ then the limit becomes

$$\lim_{(x,y)\to(0,0)} \frac{x^2 \sin y}{x^3 + 2y^3} = \lim_{x\to 0} \frac{0}{x^3} = 0.$$

From this, we might suspect that the limit actually exists. However, we need to make sure that this approaches 0 along *all possible paths* to $(0, 0)$. Here's another path we can try: $y = x$. If we do this, then the limit becomes

$$\lim_{(x,y)\to(0,0)} \frac{x^2 \sin y}{x^3 + 2y^3} = \lim_{x\to 0} \frac{x^2 \sin x}{3x^3} = \frac{1}{3}.$$

Since the value that $\frac{x^2 \sin y}{x^3+2y^3}$ approaches as $(x, y) \to (0, 0)$ depends on how (x, y) approaches $(0, 0)$, we can say that the limit does not exist. □

In general, showing limits exist for functions of several variables can be tricky. However, occasionally we can use some tricks to make these problems more manageable.

Example 12.2.3 Showing a limit exists. Find $\lim_{(x,y)\to(0,0)}(x^2+y^2)\ln(x^2+y^2)$.

Solution. Since the inside of the limit depends on $x^2 + y^2$, this suggests that switching to polar coordinates might be beneficial. So let's try that!

$$\lim_{(x,y)\to(0,0)} (x^2 + y^2)\ln(x^2 + y^2) = \lim_{(r,\theta)\to(0,0)} r^2 \ln r^2.$$

Here's why this helps us. If we look at our new limit, we see that it doesn't depend on θ at all. In other words, *it doesn't depend on the direction* we approach the origin from, just the distance from the origin r. So

$$\begin{aligned}\lim_{(r,\theta)\to(0,0)} r^2 \ln r^2 &= \lim_{r\to 0^+} 2r^3 \ln r \\ &= 0\end{aligned}$$

and so the original limit also equals 0. □

Continuous Functions

Now that we have the notion of limit, we can define continuous functions of two variables.

Definition 12.2.4 Continuous Function of Two Variables. Let $f(x, y)$ be a function of two variables. We say that f is **continuous** at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

If f is continuous at all points in a set D, we say that f is continuous on D. ◊

We won't typically use the definition of continuity in this course. We'll just make of several common types of continuous functions:

1. Polynomials in two variables, such as $x - 3x^4y^5$, are continuous on $\mathbb{R}^2$.
2. Ratios of polynomials (i.e. **rational functions**), such as $\frac{y^3-x}{1+x^{20}y^5}$, are continuous on their domains.
3. Compositions of the form $g(f(x,y))$, where $g(x)$ and $f(x,y)$ are continuous, are continuous wherever they are defined.

Example 12.2.5 Limit of a continuous function. Let $f(x,y) = \tan \frac{3x^2-5y}{1-xy}$. Find

$$\lim_{(x,y)\to(3,4)} f(x,y).$$

Solution. First, note that this is a composition of continuous functions, so $f(x,y)$ should be continuous wherever it's defined. Since this is defined at $(3,4)$, we can say that f is continuous there as well and that

$$\lim_{(x,y)\to(3,4)} f(x,y) = f(3,4) = \frac{7}{-11}.$$

□

SUGGESTED PROBLEMS: 3, 11, 19, 23, 31

12.3 Partial Derivatives

Now that we know how to take limits for functions of two variables, we can define a notion of derivative for these functions as well, which we'll call the **partial derivative**. The main idea behind the partial derivative is to measure how a function $f(x,y)$ changes as a single variable changes. Although the definition given is only for functions of two variables, it also extends to functions of many variables.

Definition 12.3.1 Partial Derivatives. Let $f(x,y)$ be a function of two variables. The **partial derivative of f with respect to x** is the function

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

assuming that this limit exists. Similarly, the **partial derivative of f with respect to y** is the function

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

assuming that this limit exists. ◊

When computing partial derivatives with respect to x, you just need to differentiate $f(x,y)$ treating it as a function of x alone and considering y as a constant. Similarly, if you want to compute f_y you differentiate with respect to y and treat x as constant.

Example 12.3.2 Let $f(x,y) = ye^{x-y} - x^3 \tan y^4$. Find $\dfrac{\partial f}{\partial y}$.

Solution. Since we need to find $\dfrac{\partial f}{\partial y}$, we differentiate $f(x,y)$ with respect to

y and treat x as a constant:

$$\begin{aligned}\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y}\left[ye^{x-y} - x^3 \tan y^4\right] \\ &= e^{x-y} - ye^{x-y} - 4x^3y^3 \sec^2 y^4\end{aligned}$$

□

Geometrically, $f_x(a,b)$ represents the rate of change of the surface $z = f(x,y)$ at the point (a,b) in the x direction, and a similar statement is true for $f_y(a,b)$.

Example 12.3.3 Partial derivatives on the unit sphere. Let $z = \sqrt{1-x^2-y^2}$. Where is $\dfrac{\partial z}{\partial y}$ equal to 0?

Solution. We can find this algebraically, but we'll try to answer this geometrically instead. First, note that $z = \sqrt{1-x^2-y^2}$ is actually the top half of the unit sphere. If we're trying to find where $\dfrac{\partial z}{\partial y}$ is zero, then we need to find where this surface is "flat" when moving in the y direction. If we think about this a bit, this should only occur when $y = 0$, i.e. along the x-axis. At any other location on the unit sphere, moving in the y-direction on the unit sphere requires going uphill or downhill, which means $z_y \neq 0$ at these locations. □

We can also find higher partial derivatives. For example, the second partial derivatives of a function $f(x,y)$ are given by

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = (f_x)_x \\ \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = (f_y)_y \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = (f_y)_x \\ \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = (f_x)_y\end{aligned}$$

The last two derivatives are examples of **mixed derivatives**.

Example 12.3.4 Mixed partials. Let $f(x,y) = \sin^2(mx+ny)$ where m, n are constants. Find the mixed partials f_{xy} and f_{yx}.

Solution. We have

$$\begin{aligned}f_{xy} &= (f_x)_y \\ &= (2m\cos(mx+ny)\sin(mx+ny))_y \\ &= -2mn\sin^2(mx+ny) + 2mn\cos^2(mx+ny) \\ &= 2mn\cos(2mx+2ny)\end{aligned}$$

as well as

$$\begin{aligned}f_{yx} &= (f_y)_x \\ &= (2n\cos(mx+ny)\sin(mx+ny))_x \\ &= -2mn\sin^2(mx+ny) + 2mn\cos^2(mx+ny) \\ &= 2mn\cos(2mx+2ny)\end{aligned}$$

□

In the last example the mixed partial were equal, but this is not always true. However, if $f(x, y)$ is a function with "nice" mixed partial derivatives, then its mixed partials will be equal.

Theorem 12.3.5 Clairaut's Theorem. *Suppose that $f(x, y)$ is defined on a disk D that contains the point (a, b). If f_{xy} and f_{yx} are both continuous on D, then*

$$f_{xy}(a, b) = f_{yx}(a, b).$$

In other words, the mixed partials are equal to each other wherever they happen to be continuous.

Partial derivatives are important for their use in **partial differential equations (PDEs)**, which are equations that involve partial derivatives of an unknown function f.

Example 12.3.6 The Wave Equation. One of the most important PDEs is the **wave equation**

$$u_{tt} = c^2 u_{xx},$$

which is used to model the displacement of a string $u(x, t)$ at position x and time t. Determine if

$$u(x, t) = \sin(x - 3t) + \ln(x + 3t)$$

is a solution of the wave equation.

Solution. We need to plug u into the PDE and check that it satisfies the PDE. Since

$$\begin{aligned} u_{tt} &= -9\cos(x - 3t) - \frac{9}{(x + 3t)^2} \\ u_{xx} &= -\sin(x - 3t) - \frac{1}{(x + 3t)^2} \end{aligned}$$

it follows that $u_{tt} = 9u_{xx}$. So $u(x, t)$ satisfies the wave equation with $c = 3$, and so is a solution of the wave equation. □

SUGGESTED PROBLEMS: 1, 3, 7, 11, 19, 31, 45, 49, 51, 53, 61

12.4 Tangent Planes and Linear Approximations

In Calculus I, derivatives are used to find linear approximations to functions of the form $f(x)$. We can use partial derivatives to do the same for functions with several independent variables.

Tangent Planes

Recall that if a curve $y = f(x)$ is differentiable at a point x_0, then it has a tangent line passing through $(x_0, f(x_0))$. The tangent line can be viewed as a linear approximation of the curve $y = f(x)$ near the point x_0. We can apply a similar ideas to surfaces $z = f(x, y)$. It turns out that if we have a surface $z = f(x, y)$ and a point $(x_0, y_0, f(x_0, y_0))$ on the surface, then every tangent vector at this point is contained in a single plane called the **tangent plane**.

We'd like to find the equation of this plane. First, recall that every plane can be described by an equation of the form

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

We can view x_0, y_0 and $z_0 = f(x_0, y_0)$ as given, so we just need to find a, b, c. If we assume that $c \neq 0$, then we can rewrite this equation to obtain

$$z - z_0 = A(x - x_0) + B(y - y_0)$$

where $A = -\frac{a}{c}$ and $B = -\frac{b}{c}$. If we set $y = y_0$, then we have $z - z_0 = A(x - x_0)$. This is the equation of a line tangent to the surface and parallel to the x-axis, and so the slope of this line must be

$$A = f_x(x_0, y_0)$$

since the slope of a tangent line in the x-direction gives the rate of change in the x-direction. Similarly,

$$B = f_y(x_0, y_0).$$

Putting all of this together gives the following theorem.

Theorem 12.4.1 Tangent Planes to Surfaces. *Let $z = f(x, y)$ be a surface and suppose that f has continuous partial derivatives at the point (x_0, y_0). Then the tangent plane to the surface $z = f(x, y)$ at the point (x_0, y_0) is given by*

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

where $z_0 = f(x_0, y_0)$.

Example 12.4.2 Approximations by tangent planes. Find the tangent plane to $f(x, y) = xe^{xy}$ at the point $(2, 0, 2)$. Use this to approximate $f(1.9, 0)$

Solution. The equation of the tangent plane is given by

$$z - 2 = (x - 2) + 4(y - 0)$$

which we can rewrite as

$$z = x + 4y.$$

So at $(1.9, 0)$, we should have

$$f(1.9, 0) \approx 1.9 + 4(0) = 1.9.$$

□

Linear Approximations

Example 12.4.2 shows that we can use tangent planes to approximate complicated functions. This leads us to the idea of a linear approximation, or **linearization** of a function of the form $f(x, y)$.

Definition 12.4.3 Linearization. Let $f(x, y)$ be a function for which f_x and f_y are both continuous at a point (x_0, y_0). Then the linearization of $f(x, y)$ at (x_0, y_0) is the function $L(x, y)$ given by

$$L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

◊

For well-behaved functions (i.e. functions that have continuous partial derivatives), $L(x, y) \approx f(x, y)$ if (x, y) is close to the point (x_0, y_0).

Example 12.4.4 Linearization of an exponential and a sinusoid. Let $u(x, t) = 30e^{-t} \sin x$. Find the linearization of u at the point $(0, \frac{\pi}{2})$.

Solution. By Definition 12.4.3, the linearization is given by

$$L(x,y) = 30 - 30(t-0) = 30(1-t).$$

□

Definition 12.4.3 can also be extended to functions with more than two variables.

Example 12.4.5 Linearization in three variables. Let $f(x,y,z) = xze^{-y^2-z^2}$. Find the linearization $L(x,y,z)$ at the point $(-3, \sqrt{\ln 2}, -\sqrt{\ln 3})$.

Solution. The formula we need to use now is

$$L(x,y,z) = f_x(-3, \sqrt{\ln 2}, -\sqrt{\ln 3})(x+3) + f_y(-3, \sqrt{\ln 2}, -\sqrt{\ln 3})(y-\sqrt{\ln 2}) + f_z(-3, \sqrt{\ln 2}, -\sqrt{\ln 3})(z+\sqrt{\ln$$

□

SUGGESTED PROBLEMS: 1, 15, 19

12.5 The Chain Rule in Several Variables

Recall that from Calc I, we know that

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

This is just the the chain rule, which tells us how to differentiate functions that depend on other functions. Or to think of it another way, it tells us how to differentiate *variables* like f that depend on other variables like g and x. What we'd like to do now is to extend the chain rule to more complicated situations. We'll start with the following formula.

Theorem 12.5.1 Chain Rule for One Independent Variable. *Let $z = f(x,y)$ be differentiable (so the partial derivatives f_x and f_y exist). If x and y both depend on t (i.e. $x = x(t)$ and $y = y(t)$), then*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

See Figure 12.5.2.

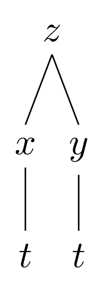


Figure 12.5.2 The chain rule for one independent variable.

Example 12.5.3 Using the multivariable chain rule. Let $f(x,y,z) = \ln\sqrt{x^2+y^2+z^2}$ where

$$x = \sin t, \quad y = \cos t \quad \text{and } z = \tan t.$$

Find $\left.\frac{df}{dt}\right]_{t=0}$.

Solution. First, let's try to simplify f. If we do so, we get

$$f(x,y,z) = \frac{1}{2}\ln(x^2+y^2+z^2).$$

Now, to find $\dfrac{df}{dt}$ we need to use the chain rule, which gives

$$\begin{aligned}\frac{df}{dt} &= f_x x'(t) + f_y y'(t) + f_z z'(t)\\ &= \frac{x}{x^2+y^2+z^2}\cos t - \frac{y}{x^2+y^2+z^2}\sin t + \frac{z}{x^2+y^2+z^2}\sec^2 t\\ &= \frac{\sin t\cos t}{\sec^2 t} - \frac{\cos t\sin t}{\sec^2 t} + \frac{\tan t}{\sec^2 t}\\ &= \cos t\sin t\\ &= \frac{1}{2}\sin 2t\end{aligned}$$

So

$$\left.\frac{df}{dt}\right]_{t=0} = 0.$$

□

If we have more than one independent variable, then the situation becomes a little bit more complicated.

Theorem 12.5.4 Chain Rule with Two Independent Variables. *Let $f(x,y)$ be differentiable, and suppose that $x = x(s,t)$ and $y = y(s,t)$. Then*

$$\begin{aligned}\frac{\partial f}{\partial s} &= f_x x_s + f_y y_s\\ \frac{\partial f}{\partial t} &= f_x x_t + f_y y_t\end{aligned}$$

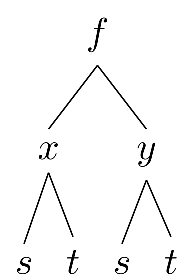


Figure 12.5.5 The chain rule for two independent variables.

Example 12.5.6 Chain rule with two independent variables. Let $f(x,y,z) = x^2 - \sin^{-1}(z-y)$, where

$$x = e^t, \quad y = s+t \quad \text{and } z = t^2 - st.$$

Find $\dfrac{\partial f}{\partial s}$.

Solution. A good way to tackle these problems is to set up a tree diagram like Figure 12.5.5 to list the dependencies between all of the variables. For $f(x,y,z)$, we get

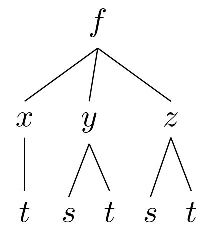


Figure 12.5.7 Tree diagram for $f(x, y, z)$.

So we can say that

$$\begin{aligned}
\frac{\partial f}{\partial s} &= f_y y_s + f_z z_s \\
&= \frac{1}{\sqrt{1-(z-y)^2}} - \frac{1}{\sqrt{1-(z-y)^2}}(-t) \\
&= \frac{1+t}{\sqrt{1-(z-y)^2}} \\
&= \frac{1+t}{\sqrt{1-(t^2-st-s-t)^2}}
\end{aligned}$$

□

Example 12.5.8 Chain rule from a tree diagram. Let $R = f(x, y, z, t)$, where $x = x(u, v, w)$, $y = y(u, v)$, $z = z(v)$ and $t = t(v, w)$. Assume that all of these functions are differentiable. Find a formula for $\dfrac{\partial R}{\partial w}$.

Solution. We can set up a tree diagram as above to find the right formula for R_w. If we do so, we obtain

$$R_w = R_x x_w + R_t t_w.$$

□

SUGGESTED PROBLEMS: 1, 7, 11, 15, 21

12.6 Directional Derivatives and Gradients

For a differentiable function $f(x, y)$, the partial derivatives f_x and f_y give the rates of change of f in the directions of the x-axis and y-axis, respectively. To put this another way, f_x gives the rate of change of f in the direction of the vector $\mathbf{i}$, and f_y gives the rate of change in the direction of the vector $\mathbf{j}$. But we have infinitely many directions to work with. How can we find the rate of change of f in *any* direction? We'll look at two tools to help us with this: directional derivatives and gradients.

The Directional Derivative

Suppose we have a function $f(x, y)$ defined at a point (x_0, y_0). We want to find how quickly $f(x, y)$ is changing in the direction of some unit vector $\mathbf{u} = \langle a, b\rangle$. We can estimate this by moving from the point (x_0, y_0) by some small increment $h\mathbf{u}$, and evaluating f at this new point, which is just $(x_0 + ha, y_0 + hb)$. So the approximate rate of change in f in the direction of $\mathbf{u}$ is given by

$$\frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

As $h \to 0$, this should become exact. This leads us to our next definition.

Definition 12.6.1 Directional Derivative. The **directional derivative** of the function $f(x, y)$ at the point (x_0, y_0) in the direction of the vector $\mathbf{u} = \langle a, b\rangle$ is given by

$$D_{\mathbf{u}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

assuming this limit exists. As we mentioned above, $D_{\mathbf{i}} f = f_x$ and $D_{\mathbf{j}} f = f_y$. ♢

We have a definition for the directional derivative, but we don't want to have to use this all the time. So we'd like to get a more practical formula.

Theorem 12.6.2 Computing the Directional Derivative. *Let $f(x, y)$ be a function whose partial derivatives f_x and f_y exist and are continuous. Let $\mathbf{u} = \langle a, b\rangle$ be a unit vector. Then $D_{\mathbf{u}} f(x, y)$ exists, and is given by*

$$D_{\mathbf{u}} f(x, y) = f_x(x, y)a + f_y(x, y)b.$$

Proof. First, set

$$g(h) = f(x_0 + ha, y_0 + hb)$$

Then by the Calculus I definition of the derivative,

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h} = D_{\mathbf{u}} f(x_0, y_0).$$

On the other hand, if we set $x = x_0 + ha$ and $y = y_0 + hb$, then $g(h) = f(x, y)$. We can then find $g'(h)$ using the chain rule from Section 3.4 to get

$$g'(h) = f_x \frac{dx}{dh} + f_y \frac{dy}{dh} = f_x(x, y)a + f_y(x, y)b$$

So $g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b$. Therefore

$$D_{\mathbf{u}} f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

■

The formula in Theorem 12.6.2 extends to functions such as $f(x, y, z)$ as well. We just need to include more partial derivatives.

Example 12.6.3 Computing a directional derivative. Let $f(r, s) = \frac{r}{r^2+s^2}$. Find the directional derivative of f in the direction of $\mathbf{v} = 3\mathbf{i} + 5\mathbf{j}$ at the point $(1, 2)$.

Solution. We can use Theorem 12.6.2 to find the directional derivative. However, there's something we need to watch out for here. Both Definition 12.6.1 and Theorem 12.6.2 require using a *unit vector*, but our vector $\mathbf{v}$ isn't a unit vector at all. So we need to **normalize v** to get a unit vector $\mathbf{u}$ that is parallel to $\mathbf{v}$:

$$\mathbf{u} = \frac{\mathbf{v}}{|\mathbf{v}|} = \left\langle \frac{3}{\sqrt{34}}, \frac{5}{\sqrt{34}} \right\rangle$$

Since

$$f_r = \frac{s^2 - r^2}{(r^2 + s^2)^2}$$
$$f_s = -\frac{2rs}{(r^2 + s^2)^2}$$

we see that the directional derivative we need is given by

$$\begin{aligned} D_{\mathbf{u}}f(1,2) &= f_r(1,2)\frac{3}{\sqrt{34}} + f_s(1,2)\frac{5}{\sqrt{34}} \\ &= \frac{3}{25}\frac{3}{\sqrt{34}} - \frac{4}{25}\frac{5}{\sqrt{34}} \end{aligned}$$

□

Gradients

Theorem 12.6.2 shows that the partial derivatives of a function are important in computing the directional derivatives. This leads us to the concept of a **gradient**.

Definition 12.6.4 Gradient of a Function. Let $f(x, y, z)$ be a function with partial derivatives f_x, f_y, f_z. The gradient of f is the vector $\nabla f = \langle f_x, f_y, f_z\rangle$. ◊

Example 12.6.5 Computing a gradient. Let $f(x, y) = x^2 + y^2$. Compute ∇f.

Solution. We have

$$\nabla f = \langle 2x, 2y\rangle$$

□

We can restate Theorem 12.6.2 very quickly in terms of the gradient. In particular, if $\mathbf{u}$ is a vector (not necessarily a unit vector) then

$$D_{\mathbf{u}}f = \nabla f \cdot \frac{\mathbf{u}}{|\mathbf{u}|} \tag{III.1}$$

Example 12.6.6 Directional derivative using the gradient. Is the function $f(x, y, z) = xy + yz + zx$ increasing or decreasing at $P(1, -1, 3)$ and in the direction of $Q(2, 4, 5)$.

Solution. We need to compute $D_{\vec{PQ}}f(1, -1, 3)$ and check if it's positive or negative. Since $\nabla f = \langle y, z, x\rangle$ and $\vec{PQ} = \langle 1, 5, 2\rangle$, it follows that

$$\begin{aligned} D_{\vec{PQ}}f(1,-1,3) &= \nabla f(1,-1,3) \cdot \frac{\vec{PQ}}{|\vec{PQ}|} \\ &= \langle -1, 3, 1\rangle \cdot \left\langle \frac{1}{\sqrt{30}}, \frac{5}{\sqrt{30}}, \frac{2}{\sqrt{30}} \right\rangle \\ &= \frac{16}{\sqrt{30}} \end{aligned}$$

Hence f is increasing at $(1, -1, 3)$ and in the direction of $(2, 4, 5)$. □

The gradient is also useful for telling us where a function is increasing and decreasing the greatest. By (III.1), we can write

$$D_{\mathbf{u}}f = \nabla f \cdot \frac{\mathbf{u}}{|\mathbf{u}|} = |\nabla f| \cos\theta,$$

where θ is the acute angle between ∇f and $\mathbf{u}$. It follows that

$$-|\nabla f| \le D_{\mathbf{u}}f \le |\nabla f|.$$

We summarize this in the following theorem.

Theorem 12.6.7 Direction of Greatest Increase and Decrease. *Let f be a function with continuous first partial derivatives. Then ∇f points in the direction of greatest increase of f, and the rate of maximum increase of f is given by $|\nabla f|$. Similarly, $-\nabla f$ points in the direction of greatest decrease, and the rate of maximum decrease of f is given by $-|\nabla f|$.*

Example 12.6.8 Finding the direction and rate of maximum decrease. Let $f(s,t) = te^{st}$. What is the direction and rate of greatest decrease of f at the point $(0,2)$?

Solution. We need to compute $-\nabla f$, which gives

$$-\nabla f = \left\langle -(1+st)e^{st}, -ste^{st} \right\rangle$$

So the direction of greatest decrease at $(0,2)$ is given by $-\nabla f(0,2) = \langle -1, 0 \rangle$. The rate of maximum decrease in this direction is $-|\nabla f(0,2)| = -|\langle -1,0\rangle| = -1$. □

SUGGESTED PROBLEMS: 1, 3, 11, 15, 21

12.7 Extreme Values

Finding the extreme values of a function remains in important goal for functions of several independent variables.

Local Maxima and Minima

One of the most important applications of the derivative in Calculus I is in finding local maxima and minima of functions $y = f(x)$. We'd like to extend this concept to functions of multiple variables, but first we need to define what local maxima and minima actually are in this new setting.

Definition 12.7.1 Local Maxima and Minima. We say that a function $f(x,y)$ has a local maximum (respectively, local minimum) at (x_0, y_0) if $f(x,y) \leq f(x_0,y_0)$ (respectively, $f(x,y) \geq f(x_0,y_0)$) for all points (x,y) contained in some disk centered at (x_0,y_0). ♢

Just as we used derivatives in Calculus I to find maxima and minima, we can use partial derivatives for that purpose here.

Theorem 12.7.2 Derivatives at a Local Maximum or Minimum. *Suppose that $f(x,y)$ is differentiable, and has a local maximum or minimum at (x_0,y_0). Then*

$$f_x(x_0,y_0) = f_y(x_0,y_0) = 0.$$

Theorem 12.7.2 does not tell us if a differentiable function $f(x,y)$ has a local maximum or minimum at (x_0,y_0), but it does tell us where to look. We call (x_0,y_0) a **critical point** of $f(x,y)$ if $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$ or if these partial derivatives fail to exist at (x_0,y_0).

Example 12.7.3 Critical points of an exponential. Let $f(x,y) = xe^{-x^2-y^2}$. Find its critical points.

Solution. Find the critical points by solving $f_x = 0$ and $f_y = 0$. Since

$$f_x = e^{-x^2-y^2} - 2x^2e^{-x^2-y^2} = e^{-x^2-y^2}(1-2x^2)$$

then $f_x = 0$ forces $x = \pm\sqrt{\frac{1}{2}}$. Similarly, since

$$f_y = -2xye^{-x^2-y^2}$$

then $f_y = 0$ forces $x = 0$ or $y = 0$. So the only points that satisfy *both* $f_x = 0$ and $f_y = 0$ are $(\pm\sqrt{\frac{1}{2}}, 0)$. So these must be our critical points. □

We found the critical points in Example 12.7.3 without too much trouble, but these are still not guaranteed to be local maxima or minima. To determine if they are, we need to use the **second derivative test**.

Theorem 12.7.4 Second Derivative Test. *Let $f(x, y)$ be a differentiable function with continuous second partial derivatives, and let (x_0, y_0) be a critical point of f. Define D to be*

$$D(x, y) = f_{xx}f_{yy} - (f_{xy})^2$$

Then we have the following possibilities:

- *If $D(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) > 0$, then (x_0, y_0) is a local minimum.*
- *If $D(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) < 0$, then (x_0, y_0) is a local maximum.*
- *If $D(x_0, y_0) < 0$, then (x_0, y_0) is a neither a local minimum nor local maximum. In this case we call (x_0, y_0) a* ***saddle point****.*
- *If $D(x_0, y_0) = 0$, the test is inconclusive.*

The quantity D from Theorem 12.7.4 is actually a determinant:

$$D(x, y) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}$$

Example 12.7.5 Local extrema of an exponential. Classify the critical points of the function from Example 12.7.3.

Solution. We have $f(x, y) = xe^{-x^2-y^2}$, and the critical points are $(-\sqrt{\frac{1}{2}}, 0)$ and $(\sqrt{\frac{1}{2}}, 0)$. We need to find $D(x, y)$. Since $f_x = (1 - 2x^2)e^{-x^2-y^2}$, we get

$$\begin{aligned} f_{xx} &= -4xe^{-x^2-y^2} - 2x(1 - 2x^2)e^{-x^2-y^2} \\ f_{xy} &= -2y(1 - 2x^2)e^{-x^2-y^2} = f_{yx} \end{aligned}$$

and similarly

$$f_{yy} = -2xe^{-x^2-y^2} + 4xy^2e^{-x^2-y^2}$$

At $(-\sqrt{\frac{1}{2}}, 0)$, we have

$$\begin{aligned} D\left(-\sqrt{\frac{1}{2}}, 0\right) &= 4\sqrt{\frac{1}{2}}e^{-\frac{1}{2}} \cdot 2\sqrt{\frac{1}{2}}e^{-\frac{1}{2}} \\ &> 0 \end{aligned}$$

and $f_{xx}(-\sqrt{\frac{1}{2}}, 0) > 0$ as well. So this is a local minimum. At $(\sqrt{\frac{1}{2}}, 0)$, we still have $D > 0$, but now $f_{xx} < 0$. So this is a local maximum. □

Absolute Maxima and Minima

Even if a function $f(x, y)$ has a local minimum at (x_0, y_0), there's no guarantee that this is the smallest value the function can take on its domain. A similar statement is also true for local maxima, which leads to the following definition.

Definition 12.7.6 Absolute Maxima and Minima. Let $f(x, y)$ be a function defined on the set D. We call the point (x_0, y_0) an **absolute minimum** (respectively, **absolute maximum**) if $f(x, y) \geq f(x_0, y_0)$ (respectively, $f(x, y) \leq f(x_0, y_0)$) for all points (x, y) in R. ♢

As a first example, consider $f(x, y) = \sqrt{1 - x^2 - y^2}$, which gives the top half of the unit sphere. Using critical points and the second derivative test, we can show that the origin is a local maximum. However, this method is blind to the absolute minimum values of $f(x, y)$, which occur at the boundary $x^2 + y^2 = 1$, where the sphere intersects the xy-plane. So this means we can't use critical points alone to find absolute maxima and minima, in general. However, the following theorem says that they'll get us most of the way there.

Theorem 12.7.7 Extreme Value Theorem. *Let $f(x, y)$ be a continuous function defined on a closed and bounded set D.* [9] *Then f is guaranteed to have an absolute maximum and absolute minimum. Furthermore, these must occur either at a critical point or on the boundary of D.*

Example 12.7.8 Absolute values on a triangle. Let $f(x, y) = x + y - xy$. Find the absolute maxima and minima of f on the triangle with vertices $(0, 0)$, $(0, 2)$ and $(4, 0)$.

Solution. We'll start off by finding any critical points. Since $\nabla f = \langle 1 - y, 1 - x\rangle$, we see that the only critical point is $(1, 1)$. Now we need to see what happens to $f(x, y)$ on the edges of the triangle, and to do that we need to find equations for the edges. Let l_1 denote the edge from $(0, 0)$ to $(0, 2)$, l_2 the edge from $(0, 2)$ to $(4, 0)$ and l_3 the edge from $(0, 0)$ to $(4, 0)$. Then l_1 is given by the equation $x = 0$, l_2 is given by $y = -\frac{1}{2}(x - 4)$ and l_3 is given by $y = 0$. So on l_1 we have $f(x, y) = f(0, y) = y$, on l_2 we have

$$f(x, y) = f\left(x, -\frac{1}{2}(x - 4)\right) = \frac{1}{2}\left[\left(x - \frac{3}{2}\right)^2 + \frac{7}{4}\right]$$

and on l_3 we have $f(x, y) = f(x, 0) = x$. We can then set up the following table:

Table 12.7.9 Absolute values on the boundary

edge	$f(x,y)$	max	min
l_1	y	2	0
l_2	$\frac{1}{2}\left[\left(x - \frac{3}{2}\right)^2 + \frac{7}{4}\right]$	$\frac{1}{2}\left[\left(4 - \frac{3}{2}\right)^2 + \frac{7}{4}\right] = 4$	$\frac{7}{8}$
l_3	x	4	0

Finally, at the critical point $(1, 1)$ we have $f(1, 1) = 1$. So putting this all together, we can say that the absolute maximum of $f(x, y)$ on this triangle is 4, and this is reached at $(4, 0)$. Similarly, the absolute minimum is 0, and this is reached at the origin. □

Note that in Example 12.7.8, we don't actually determine if the critical point is a local maximum or minimum. We just compare the value f reaches at the critical point with its maximum and minimum values on the boundary.

Example 12.7.10 Absolute maxima and minima on a circular sector. Let $f(x, y) = xy^2$ and let

$$D = \{(x, y) : x, y \geq 0, x^2 + y^2 \leq 3\}$$

Find the absolute maxima and minima of f on D.

[9] This means that D is finite in size and that it includes its boundary.

Solution. First we'll find the critical point(s). If we compute ∇f, we get $\nabla f = \langle y^2, 2xy \rangle$. This is 0 if and only if $y = 0$, and if $y = 0$ then this forces $f(x, y) = f(x, 0) = 0$.

Now we need to determine the maximum and minimum values of $f(x, y)$ on the boundary of D. Let C_1 denote the line segment from $(0, 0)$ to $(\sqrt{3}, 0)$, C_2 the circular arc from $(\sqrt{3}, 0)$ to $(0, \sqrt{3})$ and C_3 the line segment from $(0, \sqrt{3})$ to $(0, 0)$. Then on C_1 we have $f(x, y) = 0$, on C_2 we have

$$f(x, y) = xy^2 = x(3 - x^2)$$

and on C_3 we have $f(x, y) = 0$. So the smallest value that f takes on the boundary is 0, and the largest value is 2. So the absolute maximum is 2 and the absolute minimum is 0. □

SUGGESTED PROBLEMS: 1, 3, 13, 23

12.8 Lagrange Multipliers

Theorem 12.7.4 is useful for finding local maxima and minima of a function $f(x, y)$, but it's often the case where we have an additional restriction on the variables x and y, which we interpret as an equality $g(x, y) = k$ for some constant k. Geometrically, this reduces down to finding the maximum and minimum values of $f(x, y)$ on the curve $g(x, y) = k$. This problem can be attacked algebraically using contours, as in the example below.

Example 12.8.1 Constrained Optimization Using Contours. Estimate the minimum and maximum values of $f(x, y) = x^2 + 2y^2$ subject to the constraint $g(x, y) = x^2 + y^2 = 1$.

Solution. The constraint is represented as a contour plot of $g(x, y)$. We can compare this with contours of $f(x, y)$, and any point where there two contours intersect is a point for which the constraint is satisfied.

Specify a static image with the `@preview` attribute;
Or create and provide an automatic screenshot as
`generated/preview/interactive-8-preview.png`
via the `PreTeXt-CLI` application or `pretext/pretext` script.

`www.desmos.com/calculator/j3cspl30hv`

Figure 12.8.2 An interactive contour plot of $f(x, y) = s$ and $g(x, y) = 1$.

Using these contours, we see that the maximum value of f subject to $g = 1$ appears to be 2, while the minimum value of f subject to $g = 1$ appears to be 1. □

Although the graphical approach used in Example 12.8.1 and specifically in Figure 12.8.2 is useful for estimating solutions of constrained optimization problems in two dimensions, it's neither exact nor easy to visualize in higher dimensions. We therefore want to determine an algebraic approach for solving these problems. In these cases, Theorem 12.7.4 is not as useful because there's no guarantee that the critical points of f will lie on the curve $g(x, y) = k$. For this scenario, we must use the method of **Lagrange multipliers**.

Algorithm 12.8.3 Method of Lagrange Multipliers. *Suppose that f and g are both differentiable functions, and suppose that $\nabla g \neq \mathbf{0}$ on the curve*

$g(x, y) = k$. To find the extreme values of $f(x, y)$ subject to the constraint $g(x, y) = k$, perform the following steps:

1. *Find all points (x_0, y_0) such that $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$, where λ is a constant we call the **Lagrange multiplier**.*

2. *Compute f at all points from the first step and compare values to find any maxima or minima.*

Example 12.8.4 Maximizing volume. A lidless box may be made from $12\,\mathrm{m}^2$ of material. What dimensions give the maximum volume?

Solution. If we denote the dimensions of the box by x, y, z, then the function we need to maximize is $V(x, y, z) = xyz$. However, not all dimensions are valid since we're only allowed to use $12\,\mathrm{m}^2$ of material. This means that the surface area of the box, which is given by $A(x, y, z) = xy + 2xz + 2yz$, must be equal to 12. So our constraint is that $A(x, y, z) = 12$, and because we have this constraint we'll try to solve this using Algorithm 12.8.3.

To begin, we need to solve $\nabla V = \lambda \nabla A$. Equivalently, we need to solve the system

$$\begin{aligned} V_x &= \lambda A_x \\ V_y &= \lambda A_y \\ V_z &= \lambda A_z \end{aligned}$$

which turns into

$$\begin{aligned} yz &= \lambda(y + 2z) \\ xz &= \lambda(x + 2z) \\ xy &= \lambda(2x + 2y) \end{aligned}$$

Now, we don't actually need to find a value for λ. We just need values for x, y, z. If we multiply these equations by x, y and z respectively, then we get

$$\begin{aligned} xyz &= \lambda(xy + 2xz) \\ xyz &= \lambda(xy + 2yz) \\ xyz &= \lambda(2xz + 2yz) \end{aligned}$$

or just

$$\lambda(xy + 2xz) = \lambda(xy + 2yz) = \lambda(2xz + 2yz)$$

We can also divide by λ, since there's no way λ could equal 0 and satisfy the above equations (since x, y, z are all positive). This gives us

$$xy + 2xz = xy + 2yz = 2xz + 2yz \tag{III.2}$$

The first two expressions in (III.2) simplify to $x = y$, while the first and third reduce to $x = 2z$. So

$$x = y = 2z.$$

But these variables also need to satisfy $A(x, y, z) = 12$:

$$\begin{aligned} 12 &= xy + 2xz + 2yz \\ &= 4z^2 + 4z^2 + 4z^2 \\ &= 12z^2 \end{aligned}$$

Therefore $z = \pm 1$, so we take $z = 1$ (since we're dealing with dimensions). This also means that $x = y = 2(1) = 2$. So the point $(2, 2, 1)$ is therefore our candidate for the dimensions that maximize volume subject to the constraint $A = 12$.

In order to actually verify that $V(2,2,1) = 4$ is actually the largest possible volume we can obtain, we need to show that it's actually a maximum and not a minimum. This means we need to find another point (x, y, z) that also satisfies $A = 12$, and show that it gives a smaller value for V. One such point is just $(-2,-2,-1)$, and $V(-2,-2,-1) = -4 < V(2,2,1)$. So 4 is actually a maximum value and not a minimum value. □

The points we'll obtain using Lagrange multipliers will be either maxima or minima, but the method itself doesn't always tell us which is which. As in the last example, sometimes it's up to us to find that on our own.

Example 12.8.5 Extreme values on a paraboloid. Find the absolute maxima and minima for $f(x,y) = x^2 + 2y^2$ over this disk $x^2 + y^2 \leq 1$.

Solution. First, recall from Section 12.7 that the absolute maxima and minima of a differentiable function over a bounded set occur at either a critical point or somewhere on the boundary. So first, we need to find if $f(x, y)$ has any critical points inside the disk. If we solve $f_x = f_y = 0$, then we quickly get that $(0, 0)$ is a critical point, and hence a possible maximum or minimum value. The only other place we need to check is the boundary $x^2 + y^2 = 1$, and this is something we can use Lagrange multipliers for.

Set $g(x, y) = x^2 + y^2$. Then we need to find the extreme values of f subject to the constraint $g = 1$. The system of equations we need to solve is then

$$\begin{aligned} 2x &= \lambda(2x) & &\Rightarrow x = \lambda x \\ 4y &= \lambda(2y) & &\Rightarrow 2y = \lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

The first equation is true if $\lambda = 1$, but then this forces $y = 0$ in the second equation. If we then plug $y = 0$ into the constraint, this forces $x = \pm 1$. So the case $\lambda = 1$ gives us two points to test: $(-1, 0)$ and $(1, 0)$. On the other hand, the second equation is true if $\lambda = 2$. This then forces $x = 0$ and $y = \pm 1$, which gives us the points $(0,-1)$ and $(0, 1)$. Finally, if $\lambda \neq 1, 2$, then this forces $x = y = 0$. But there's no way to satisfy the constraint if $x = y = 0$, so we discard this possibility.

So we have five points that we need to check, which we arrange in the following table:

Table 12.8.6 Absolute values on the disk

(x, y)	$f(x, y)$
$(0, 0)$	0
$(-1, 0)$	1
$(1, 0)$	1
$(0, -1)$	2
$(0, 1)$	2

So the absolute minimum is 0, which occurs at the origin. The absolute maximum is 2, which occurs at the points $(0, \pm 1)$. □

Example 12.8.7 Maximizing the volume of a prism within an ellipsoid. A rectangular prism, centered at the origin and with sides parallel to the coordinate planes, is inscribed within the ellipsoid $x^2 + \frac{y^2}{4} + \frac{z^2}{4} = 1$. Find the dimensions of the prism that maximize the volume.

Solution. Let (x, y, z) denote the corner of the prism located in the first octant. Then we want to maximize $V(x, y, z) = (2x)(2y)(2z)$ subject to the constraint $g(x, y, z) = x^2 + \frac{y^2}{4} + \frac{z^2}{4} = 1$. So we'll use the method of Lagrange multipliers again to get the system

$$\begin{aligned} 8yz &= 2\lambda x \\ 8xz &= \frac{\lambda}{2} y \\ 8xy &= \frac{\lambda}{2} z \\ x^2 + \frac{y^2}{4} + \frac{z^2}{4} &= 1 \end{aligned}$$

We can assume that x, y, z are all positive (since we're trying to find the *maximum* volume), so we can go ahead and solve each equation for λ to get

$$\lambda = \frac{4yz}{x} = \frac{16xz}{y} = \frac{16xy}{z}$$

Setting the first two fractions equal and simplifying gives $y^2 = 4x^2$. Similarly, the second and third fractions give $y^2 = z^2$. So

$$y^2 = z^2 = 4x^2.$$

Now we plug this into our constraint to get

$$\begin{aligned} 1 &= x^2 + \frac{y^2}{4} + \frac{z^2}{4} = 1 \\ &= 3x^2 \end{aligned}$$

so $x = \pm\sqrt{\frac{1}{3}}$. Since we're assuming that (x, y, z) lies in the first octant, this means that $x > 0$, along with $y$and z. So

$$\begin{aligned} x &= \sqrt{\frac{1}{3}} \\ y &= 2\sqrt{\frac{1}{3}} \\ z &= 2\sqrt{\frac{1}{3}} \end{aligned}$$

To check that these values actually maximize $V(x, y, z)$, we can compare them with the point $(1, 0, 0)$, which also satisfies the constraint $g = 1$. Since $V(1, 0, 0) = 0$, this means that the dimensions that maximize the volume are given by

$$\frac{2}{\sqrt{3}} \times \frac{4}{\sqrt{3}} \times \frac{4}{\sqrt{3}}$$

□

SUGGESTED PROBLEMS: 1-11 odd, 17, 19

Chapter 13

Multiple Integrals

From Calculus I, we know how to find the area under the graph $y = f(x)$ from $x = a$ to $x = b$: it's just

$$\int_a^b f(x)\,dx.$$

What we want to do now is to take the notion of integration and extend it to higher dimensions. As a starting motivation, we'd like to develop the concept of an integral of a function $f(x, y)$ of two independent variables in order to find the volume under the surface $z = f(x, y)$.

13.1 Double Integrals over Rectangles

We want to try to find the volume under the surface $z = f(x, y)$ and above some rectangle R in the xy-plane. If $f(x, y)$ is a constant function, then this is easy: the resulting surface and the xy-plane then form a rectangular prism. If $f(x, y)$ is more complicated, then we might not have a nice volume formula to use. However, we can steal an idea from Calculus I and try to approximate the volume by using simpler shapes; in this case, rectangular prisms.

Riemann Sums

Given a continuous function $f(x, y)$ defined over the rectangle

$$R = \{(x, y) : a \le x \le b, c \le y \le d\}$$

to approximate the volume under f and above R we'll divide the rectangle up into smaller rectangles. In particular, let's divide the intervals $[a, b]$ and $[c, d]$ into subintervals $[x_i, x_{i+1}], [y_j, y_{j+1}]$, where

$$\begin{aligned}
a = x_0 < x_1 < x_2 < \cdots < x_m = b \\
c = y_0 < y_1 < y_2 < \cdots < y_n = d \\
\Delta x = x_{i+1} - x_i \\
\Delta y = y_{j+1} - y_j
\end{aligned}$$

If we let R_{ij} denote the subrectangle given by

$$R_{ij} = \{(x, y) : x_i \le x \le x_{i+1}, y_j \le y \le y_{j+1}\}$$

then the volume under $f(x,y)$ and above R_{ij} is about equal to

$$f(x_i, y_j)\Delta x \Delta y$$

Hence the volume under $f(x,y)$ and above R should be about equal to

$$\sum_{i=0}^{m}\sum_{j=0}^{n} f(x_i, y_j)\Delta x \Delta y$$

If we replace $\Delta x \Delta y$ with ΔA and send $\Delta A \to 0$, this approximation becomes exact. This gives us the following definition.

Definition 13.1.1 Double Integral Over a Rectangle. Let $f(x,y)$ be a function defined over the rectangle R. Then the **double integral** of $f(x,y)$ over R is defined to be

$$\iint_R f(x,y)\,dA = \lim_{\Delta A \to 0}\sum_{i=0}^{m}\sum_{j=0}^{n} f(x_i, y_j)\Delta A$$

If the limit exists, we say that f is **integrable** over R. ◊

The geometric intuition behind the double integral is that it represents a volume under a surface. Negative values for a double integral indicate that a surface tends to lie below the xy-plane, while positive values indicate that the surface tends to lie above the xy-plane. Furthermore, continuous functions are integrable over any rectangle R.

Example 13.1.2 Double integrals by volume. Let $f(x,y) = 5 - x$ and let

$$R = \{(x,y) : 0 \le x \le 5, 0 \le y \le 3\}$$

Find $\iint_R f(x,y)\,dA$.

Solution. If we graph $f(x,y)$, we get a triangular cylinder running along the y axis. The volume of this cylinder over R is just the area of the triangular "base" times the "height," or just

$$\frac{1}{2}5 \cdot 5 \cdot 3 = \frac{75}{2}$$

So

$$\iint_R (5-x)\,dA = \frac{75}{2}$$

□

Fubini's Theorem

Now that we have a definition for the double integral, we want to find a better way of computing it. Thankfully, we do have such a method: Fubini's Theorem.

Theorem 13.1.3 Fubini's Theorem. *Let $f(x,y)$ be continuous on the rectangle*

$$R = \{(x,y) : a \le x \le b, c \le y \le d\}$$

Then

$$\iint_R f(x,y)\,dA = \int_a^b \left[\int_c^d f(x,y)\,dy\right] dx = \int_c^d \left[\int_a^b f(x,y)\,dx\right] dy$$

Example 13.1.4 Double integrals by Fubini's Theorem. Let $f(x, y) = 5 - x$ and let

$$R = \{(x, y) : 0 \le x \le 5, 0 \le y \le 3\}$$

Find $\iint_R f(x, y)\, dA$.

Solution. We found this previously in Example 13.1.2, so we'll try it again using Fubini's Theorem and see if we get the same answer. By Theorem 13.1.3, we have

$$\begin{aligned}\iint_R f(x, y)\, dA &= \int_0^3 \left[\int_0^5 (5 - x)\, dx\right] dy \\ &= \int_0^3 \left[5x - \frac{x^2}{2}\right]_{x=0}^5 dy \\ &= \int_0^3 \frac{25}{2}\, dy \\ &= \frac{75}{2}\end{aligned}$$

□

The double integral, like the Calculus I integral, derivative and partial derivatives, is *linear*. This means that you can break it up over addition and pull constants outside of it. In other words, if f and g are integrable and if c is a constant, then

$$\iint_R (f(x, y) + g(x, y))\, dA = \iint_R f(x, y)\, dA + \iint_R g(x, y)\, dA$$
$$\iint_R cf(x, y)\, dA = c \iint_R f(x, y)\, dA$$

Example 13.1.5 Double integral of a logarithm. Find

$$\int_1^3 \int_1^3 \frac{\ln(xy)}{xy}\, dy\, dx$$

Solution. First, note that

$$\int_1^3 \int_1^3 \frac{\ln(xy)}{xy}\, dy\, dx = \int_1^3 \int_1^3 \frac{\ln(x)}{xy}\, dy\, dx + \int_1^3 \int_1^3 \frac{\ln(y)}{xy}\, dy\, dx$$

To find the first double integral on the right, first we integrate with respect to y to get

$$\int_1^3 \int_1^3 \frac{\ln(x)}{xy}\, dy\, dx = \int_1^3 \frac{\ln x}{x} \ln y\Big|_{y=1}^3 dx = \ln 3 \int_1^3 \frac{\ln x}{x}\, dx$$

Now we can use u-sub with $u = \ln x$ and $du = \frac{1}{x}\, dx$ to get

$$\begin{aligned}\ln 3 \int_1^3 \frac{\ln(x)}{x}\, dx &= \ln 3 \frac{u^2}{2}\Big|_{u=0}^{\ln 3} \\ &= \frac{1}{2}(\ln 3)^3\end{aligned}$$

The same exact process shows that $\int_1^3 \frac{\ln y}{xy}\, dy\, dx$ must also equal $\frac{1}{2}(\ln 3)^3$. Therefore

$$\int_1^3 \int_1^3 \frac{\ln(xy)}{xy}\, dy\, dx = (\ln 3)^3$$

□

SUGGESTED PROBLEMS: 7, 9, 15, 19

13.2 Double Integrals over General Regions

Section 13.1 shows how to define the double integral over a rectangle R. Now we want to extend it to more general regions. We'll be skipping an issue with how to actually define such an integral in terms of Riemann sums, but the main idea is to take a function defined over some general region D and then extend it to cover a rectangular region R. Then we can just use Definition 13.1.1 to define this new integral as well.

Our primary tool for computing double integrals will still be Theorem 13.1.3. The main difficulty we'll encounter (aside from the usual integration issues) is how to properly set up limits for x and y for some region D.

Example 13.2.1 Integrating over a region defined by a parabola. Let $f(x, y) = \frac{y}{x^5+1}$ and let D be the region under the parabola $y = x^2$, above the x-axis and between $x = 0$ and $x = 1$. Find $\iint_D f(x, y)\, dA$.

Solution. The first step is to determine limits for x and y that describe this region. One possible description is the following:

$$D = \{(x, y) : 0 \le y \le x^2, 0 \le x \le 1\}$$

If we use this, we can write

$$\begin{aligned}\iint_D \frac{y}{x^5+1}\, dA &= \int_0^1 \int_0^{x^2} \frac{y}{x^5+1}\, dy\, dx \\ &= \frac{1}{2}\int_0^1 \frac{1}{x^5+1} y^2 \Big|_{y=0}^{x^2} dx \\ &= \frac{1}{2}\int_0^1 \frac{x^4}{x^5+1}\, dx \\ &= \frac{1}{10}\int_1^2 \frac{1}{u}\, du \\ &= \frac{1}{10}\ln 2\end{aligned}$$

We also could have described D as the set of all points

$$D = \{(x, y) : \sqrt{y} \le x \le 1, 0 \le y \le 1\}$$

If we do this instead, we get

$$\iint_D \frac{y}{x^5+1}\, dA = \int_0^1 \int_{\sqrt{y}}^1 \frac{y}{x^5+1}\, dx\, dy$$

By Theorem 13.1.3, this is guaranteed to equal $\frac{1}{10}\ln 2$. However, the actual process of computing the double integral using this choice for the limits is much more difficult, since we would need to integrate $\frac{1}{x^5+1}$ with respect to x. □

There are several important things we can take away from the above example:

- Much of the difficulty in computing double integrals lies in finding limits that describe the region you're integrating over. In general, it's a good idea to sketch the region you're integrating over in order to figure out what your limits should be.

- There can be multiple ways to describe a single region. This leads to multiple ways of setting up limits for double integrals over this region. If the integrand is continuous, then Theorem 13.1.3 guarantees that these different methods will all produce the same value.
- It's important to choose limits in such a way as to make computing the double integral more manageable.
- When finding limits for a double integral, the outermost limits *must be constant* since the double integral must eventually simplify to a constant. In other words, we don't really have a notion of an "indefinite" double integral.

Example 13.2.2 Reversing the order of integration. Compute $\int_0^{\sqrt{\pi}} \int_y^{\sqrt{\pi}} \cos x^2\, dx\, dy$.

Solution. If we try to integrate with respect to x right away, then we're stuck: $\cos x^2$ has no elementary antiderivative, which for us means that we can only integrate it using power series. However, it is continuous, so Theorem 13.1.3 tells us that if we can integrate it with respect to y first and then x (i.e. replace $dx\, dy$ with $dy\, dx$), we'll still get the same answer. And integrating $\cos x^2$ with respect to y is actually very easy. This is called **reversing the order of integration**.

In order to change the order of integration, we need to change the limits. The region we're integrating over is the region

$$D = \{(x, y) : y \le x \le \sqrt{\pi}, 0 \le y \le \sqrt{\pi}\}$$

If we sketch this, we see that this is just the region below the line $y = x$, above the x-axis and bounded from $x = 0$ to $x = \sqrt{\pi}$. So we can also write

$$D = \{(x, y) : 0 \le y \le x, 0 \le x \le \sqrt{\pi}\}$$

Therefore

$$\begin{aligned}
\int_0^{\sqrt{\pi}} \int_y^{\sqrt{\pi}} \cos x^2\, dx\, dy &= \int_0^{\sqrt{\pi}} \int_0^{x} \cos x^2\, dy\, dx \\
&= \int_0^{\sqrt{\pi}} y \cos x^2 \Big|_{y=0}^{x} dx \\
&= \int_0^{\sqrt{\pi}} x \cos x^2\, dx \\
&= \frac{1}{2} \int_0^{\pi} \cos u\, du \\
&= 0
\end{aligned}$$

□

Sometimes, the order of integration can be chosen to make the limits of integration easier to set up.

Example 13.2.3 Integrating over a Triangle. Integrate $f(x, y) = x^2 + y^2$ over the triangle T with vertices $(0, 0), (1, 1)$ and $(0, 2)$.

Solution. If we wanted to integrate with respect to x first, we would have to break our double integral up into two different double integrals. This is because the limits for x change halfway up the triangle. So it'd be better for us to integrate with respect to y first. Since

$$T = \{(x, y) : x \leq y \leq 2 - x, 0 \leq x \leq 1$$

we have

$$\begin{aligned}
\iint_T (x^2 + y^2)\, dA &= \int_0^1 \int_x^{2-x} (x^2 + y^2)\, dy\, dx \\
&= \int_0^1 \left[x^2 y + \frac{y^3}{3} \right]_{y=x}^{2-x} dx \\
&= \int_0^1 \left[x^2(2 - 2x) + \frac{1}{3}[(2 - x)^3 - x^3] \right] dx \\
&= \left[\frac{2x^3}{3} - \frac{1}{2}x^4 - \frac{(2 - x)^4}{12} - \frac{x^4}{4} \right]_0^1 \\
&= \frac{2}{3} - \frac{1}{2} - \frac{1}{12} - \frac{1}{4} + \frac{16}{12} \\
&= \frac{7}{6}
\end{aligned}$$

□

Besides finding volumes, we can also use double integrals to find areas. If D is some region in the xy-plane, then $\iint_D dA$ represents the volume under the surface $z = 1$ and above D. This is a solid with the constant height of 1, so the volume should be equal to the area of D times 1. That is, $\iint_D dA$ is equal to the area of D.

SUGGESTED PROBLEMS: 1, 5, 11, 13, 17, 21

13.3 Double Integrals in Polar Coordinates

Recall that the double integral was defined by first setting up a rectangular grid. The reason we used a rectangular grid was because we were working in Cartesian coordinates, so this made the most sense. If we're dealing with a circular region of integration, then using Cartesian coordinates is very awkward. However, polar coordinates from Section 10.3 work very nicely with circular regions. So we want to find out how to set up double integrals using polar coordinates.

If we're given a function $f(x, y)$, then it's not too hard to convert this to the polar form $f(r, \theta)$. Just replace x with $r \cos \theta$ and y with $r \sin \theta$. The tricky part with setting up double integrals in polar coordinates is how to deal with the **area element** dA, which in Cartesian coordinates is just $dx\, dy$ or $dy\, dx$. To figure out what dA should be in polar coordinates, i.e. in terms of r and θ, consider the following "polar rectangle":

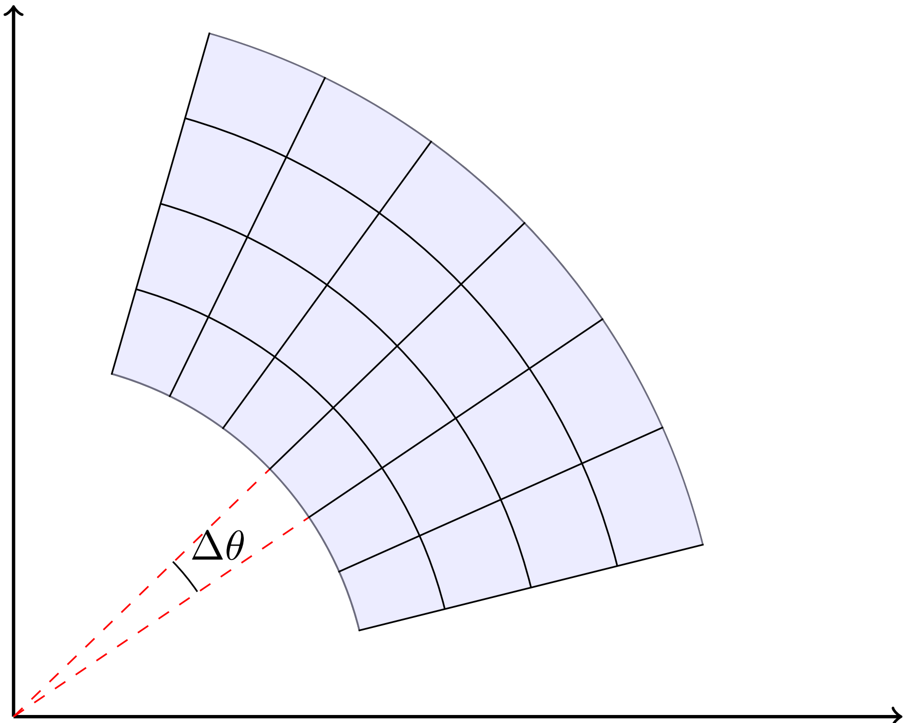


Figure 13.3.1 A polar grid.

Let ΔA represent the area of one of these sectors. If we let r denote the distance from the origin to one sector, Δr the length of a sector and $\Delta\theta$ the angle spanned by a sector, then we can say that

$$\begin{aligned}\Delta A &\approx \frac{1}{2}[r\cos(\Delta\theta)][r\sin(\Delta\theta)] - \frac{1}{2}[(r-\Delta r)\cos(\Delta\theta)][(r-\Delta r)\sin(\Delta\theta)]\\ &= r\Delta r\cos(\Delta\theta)\sin(\Delta\theta) - \frac{1}{2}(\Delta r)^2\cos(\Delta\theta)\sin(\Delta\theta)\end{aligned}$$

If we assume that Δr and $\Delta\theta$ are both small (which means the polar grid in Figure 13.3.1 is very fine), then

$$\begin{aligned}\cos(\Delta\theta) &\approx 1\\ \sin(\Delta\theta) &\approx \Delta\theta\\ (\Delta r)^2 &\approx 0\end{aligned}$$

So

$$\Delta A \approx r\Delta r\Delta\theta.$$

As Δr and $\Delta\theta$ approach 0, this becomes more exact, and we get

$$dA = r\,dr\,d\theta.$$

Theorem 13.3.2 Double Integrals in Polar Coordinates. *Let $f(x,y)$ be a continuous function. Then*

$$\iint_D f(x,y)\,dA = \iint_D f(r\cos\theta, r\sin\theta)r\,dr\,d\theta$$

and limits are chosen using polar coordinates.

Example 13.3.3 Integrating over a circular sector. Find

$$\int_{-1}^{0}\int_{-\sqrt{1-x^2}}^{0} \frac{2}{1+\sqrt{x^2+y^2}}\,dy\,dx$$

Solution. If we sketch the region of integration, we see that it is the part of the unit circle in the third quadrant. So we'll switch to polar coordinates to solve this integral:

$$\begin{aligned}\int_{-1}^{0}\int_{-\sqrt{1-x^2}}^{0} \frac{2}{1+\sqrt{x^2+y^2}}\,dy\,dx &= \int_{\pi}^{\frac{3\pi}{2}}\int_{0}^{1} \frac{2}{1+r} r\,dr\,d\theta\\ &= 2\int_{\pi}^{\frac{3\pi}{2}} \frac{r}{1+r}\,dr\,d\theta\\ &= 2\int_{\pi}^{\frac{3\pi}{2}} \Big[r-\ln(1+r)\Big]_{r=0}^{1}\,d\theta\\ &= \pi(1-\ln 2)\end{aligned}$$

□

Polar coordinates may also be used, surprisingly, to evaluate the **Gaussian integral** $\int_{-\infty}^{\infty} e^{-x^2}\,dx$.

Theorem 13.3.4 The Gaussian Integral. *We have*

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

Proof. First, let $I = \int_{-\infty}^{\infty} e^{-x^2}\,dx$. We'll show that $I^2 = \pi$. We have

$$\begin{aligned}I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2}\,dy\right)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy\\ &= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2} r\,dr\,d\theta\\ &= \int_{0}^{2\pi} \lim_{b\to\infty} \frac{-1}{2}e^{-r^2}\Big]_{r=0}^{b}\,d\theta\\ &= \int_{0}^{2\pi}\frac{1}{2}\,d\theta\\ &= \pi\end{aligned}$$

Since $I^2 = \pi$, this gives

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

■

Example 13.3.5 Volume of a sphere. Find the volume of a sphere of radius ρ.

Solution. First, we can center the sphere at the origin without loss of generality. Such a sphere is given by $x^2 + y^2 + z^2 = \rho^2$. If we solve for z, we get

$$z = \pm\sqrt{\rho^2 - x^2 - y^2}$$

Let D denote the disk of radius ρ in the xy-plane centered at the origin. Then the volume of the sphere is

$$\begin{aligned}
2\iint_D \sqrt{\rho^2 - x^2 - y^2}\,dA &= 2\int_0^{2\pi}\int_0^{\rho} \sqrt{\rho^2 - r^2}r\,dr\,d\theta \\
&= -\int_0^{2\pi}\int_{\rho^2}^{0} \sqrt{u}\,du\,d\theta \\
&= \frac{2}{3}\int_0^{2\pi} u^{3/2}\Big]_{u=0}^{\rho^2} d\theta \\
&= \frac{4\pi}{3}\rho^3
\end{aligned}$$

□

SUGGESTED PROBLEMS: 5, 13, 23

13.4 Applications of Double Integrals

Mass

Consider a thin plate in the xy-plane, say the region R. If the density of the plate $\rho(x,y)$ at (x,y) is constant, say $\rho(x,y) = C$, then the mass of the plate is just the density times the area. On the other hand, if the density is varies then this becomes more complicated, and we must use double integrals. In particular, the mass of a plate contained in the region R in the xy-plane with density $\rho(x,y)$ is given by

$$\iint_R \rho(x,y)\,dA$$

Example 13.4.1 Mass of a triangular plate. Find the mass of the plate contained in the triangular region D bounded by lines $y = x$, $y = 6$ and $2x+y = 6$, given that the density is $\rho(x,y) = x^2$.

Solution. The mass is

$$\begin{aligned}
\iint_D x^2\,A &= \int_2^6\int_{\frac{6-y}{2}}^{y} x^2\,dx\,dy \\
&= \frac{1}{3}\int_2^6 \left[\frac{y^3 - (6-y)^3}{8}\right] dy
\end{aligned}$$

□

The **moments** of the mass contained in D and with density $\rho(x,y)$ are defined as follows:

$$\begin{aligned}
M_x &= \iint_D y\rho(x,y)\,dA \\
M_y &= \iint_D x\rho(x,y)\,dA
\end{aligned}$$

If we let m denote the total mass, then we also define the **center of mass** (or **centroid**) to be the point $(\bar{x},\bar{y})$, where

$$\begin{aligned}
\bar{x} &= \frac{M_y}{m} \\
\bar{y} &= \frac{M_x}{m}
\end{aligned}$$

Example 13.4.2 Center of mass of an annulus. Find the center of mass of the plate D outside of the circle $x^2+y^2=1$ and inside the circle $x^2+y^2=4$, with density $\rho(x,y)=\frac{1}{\sqrt{x^2+y^2}}$.

Solution. The mass is given by

$$\begin{aligned}\iint_D \frac{1}{\sqrt{x^2+y^2}}\,dA &= \int_0^{2\pi}\int_1^2 \frac{r}{r}\,dr\,d\theta \\ &= 2\pi\end{aligned}$$

The moments are

$$\begin{aligned}M_x &= \iint_D \frac{y}{\sqrt{x^2+y^2}}\,dA \\ &= \int_0^{2\pi}\int_1^2 \frac{r\sin\theta}{r} r\,dr\,d\theta \\ &= 0\end{aligned}$$

and likewise $M_y=0$. So the center of mass is $(0,0)$. □

SUGGESTED PROBLEMS: 3, 5, 13

13.5 Triple Integrals

In $\mathbb{R}$, we have

$$\int_a^b f(x)\,dx = \lim_{\Delta x\to 0}\sum_{i=0}^{n} f(x_i)\Delta x.$$

This represents the area under $y=f(x)$ and over $[a,b]$. Furthermore, $\int_a^b dx$ gives the length of $[a,b]$. In $\mathbb{R}^2$, we have

$$\int_D f(x,y)\,dA = \lim_{\Delta A\to 0}\sum_{i=0}^{m}\sum_{j=0}^{n} f(x_i,y_j)\Delta A,$$

This represents the volume under $z=f(x,y)$ and above the region D, where $\Delta A=\Delta x\Delta y$. Furthermore, $\iint_D dA$ gives the area of D.

We can extend all of this to $\mathbb{R}^3$ by introducing the concept of the **triple integral**.

Definition 13.5.1 Triple Integrals over a Rectangle. Let $f(x,y,z)$ be defined on some region D in $\mathbb{R}^3$. Then the triple integral of f over D is given by

$$\iiint_D f(x,y,z)\,dV = \lim_{\Delta V\to 0}\sum_{i=0}^{l}\sum_{j=0}^{m}\sum_{k=0}^{n} f(x_i,y_j,z_k)\Delta V$$

where $\Delta V=\Delta x\Delta y\Delta z$. If the limit exists, we say that f is **integrable** on D. ◊

For a double integral in rectangular coordinates, we have $dA=dx\,dy$ or $dy\,dx$. Similarly, for a triple integral in rectangular coordinates we have six different choices for dV:

Table 13.5.2 Absolute values on the disk

$dx\,dy\,dz$	$dx\,dz\,dy$
$dy\,dx\,dz$	$dy\,dz\,dx$
$dz\,dx\,dy$	$dz\,dy\,dx$

Just as we can view dx as an infinitesimal length and dA as an infinitesimal area, dV represents an infinitesimal volume.Our main use for Definition 13.5.1 will be to recognize a triple integral ”in the wild,” but we won’t actually use it to compute integrals. For this purpose, we still use Fubini’s Theorem.

Theorem 13.5.3 Fubini’s Theorem for Triple Integrals. *Suppose $f(x, y, z)$ is a continuous function on the closed and bounded region D in $\mathbb{R}^3$. Then $\iiint_D f(x, y, z)\,dV$ can be computed as an iterated integral, and the answer does not depend on the choice of dV.*

Example 13.5.4 A triple integral over a rectangular prism. Compute $\iiint_D xyze^{-x^2-y^2}\,dV$, where

$$D = \{(x, y) : 0 \le x \le \sqrt{\ln 2}, 0 \le y \le \sqrt{\ln 4}, 0 \le z \le 1\}$$

Solution. We’ll integrate using $dV = dx\,dy\,dz$. Then we have

$$\begin{aligned}
\iiint_D xyze^{-x^2-y^2}\,dV &= \int_0^1 \int_0^{\sqrt{\ln 4}} \int_0^{\sqrt{\ln 2}} xyze^{-x^2-y^2}\,dx\,dy\,dz \\
&= \int_0^1 \int_0^{\sqrt{\ln 4}} \left[-\frac{yze^{-x^2-y^2}}{2}\right]_{x=0}^{\sqrt{\ln 2}} \\
&= -\frac{1}{2}\int_0^1 \int_0^{\sqrt{\ln 4}} yz(e^{-\ln 2-y^2} - e^{-y^2})\,dy\,dz \\
&= -\frac{e^{-\ln 2}-1}{2}\int_0^1 \int_0^{\sqrt{\ln 4}} yze^{-y^2}\,dy\,dz \\
&= \frac{-\frac{1}{2}}{4}\int_0^1 \left[ze^{-y^2}\right]_{y=0}^{\sqrt{\ln 4}}\,dz \\
&= \frac{\frac{3}{4}}{8}\int_0^1 z\,dz \\
&= \frac{3}{32}\frac{1}{2} \\
&= \frac{3}{64}
\end{aligned}$$

□

An unfortunate side effect of increasing the dimension for our integral is that we lose a little bit of geometric intuition. For instance, Example 13.5.4 is indeed calculating a ”volume,” but the volume in question is for a four dimensional region (the graph of $f(x, y, z)$ over the rectangular prism). We can only really visualize the ”base” of this region, which served as our region of integration in Example 13.5.4. Even so, the triple integral can still tell us important things about functions of three variables.

Example 13.5.5 Finding an average value. Find the average value of the function $f(x, y, z) = xyze^{-x^2-y^2}$ over the region D given in Example 13.5.4.

Solution. First, let Vol(D) denote the volume of D. Then the average value of f over D is just

$$\frac{\iiint_D xyze^{-x^2-y^2}\,dV}{\mathrm{Vol}(D)} = \frac{\frac{3}{64}}{\sqrt{\ln 2 \ln 4}} \approx .05$$

□

We can also compute triple integrals over more general regions.

Example 13.5.6 Volume using triple integrals. Find the volume of the region bounded by the cylinder $y = x^2$ and the planes $z = 3 - y$ and $z = 0$.

Solution. If we let D denote this region, then its volume is given by $\iiint_D dV$. The volume is then

$$\begin{aligned}
\iiint_D dV &= \int_{-\sqrt{3}}^{\sqrt{3}} \int_{x^2}^{3} \int_0^{3-y} dz\,dy\,dx \\
&= \int_{-\sqrt{3}}^{\sqrt{3}} \int_{x^2}^{3} (3-y)\,dy\,dx \\
&= \int_{-\sqrt{3}}^{\sqrt{3}} \left[3y - \frac{y^2}{2}\right]_{y=x^2}^{3} dx \\
&= \int_{-\sqrt{3}}^{\sqrt{3}} \left[9 - \frac{9}{2} - 3x^2 + \frac{x^2}{2}\right] dx \\
&= \frac{24\sqrt{3}}{2}
\end{aligned}$$

□

When setting up limits for triple integrals, say using $dV = dx\,dy\,dz$, then the limits on the innermost integral are typically functions of y and z, the limits on the middle integral are functions of z and the limits on the outermost integral are constant. We can also change the order of integration to make an integral more tractable.

Example 13.5.7 Changing the order of integration. Compute $\int_1^4 \int_z^{4z} \int_0^{\pi^2} \frac{\sin\sqrt{yz}}{x^{3/2}}\,dy\,dx\,dz$.

Solution. This looks awful to integrate with respect to y first, so we'll try changing the order of integration. x looks easiest, so let's try using $dV = dx\,dy\,dz$ instead. If we sketch the region, we see that the limits are actually the same, expect we just need to swap the middle and innermost integrals. So

$$\begin{aligned}
\int_1^4 \int_z^{4z} \int_0^{\pi^2} \frac{\sin\sqrt{yz}}{x^{3/2}}\,dy\,dx\,dz &= \int_1^4 \int_0^{\pi^2} \int_z^{4z} x^{-3/2} \sin\sqrt{yz}\,dx\,dy\,dz \\
&= \int_1^4 \int_0^{\pi^2} \left[-2x^{-1/2}\sin\sqrt{yz}\right]_{x=z}^{4z} dy\,dz \\
&= \int_1^4 \int_0^{\pi^2} z^{-1/2} \sin\sqrt{yz}\,dy\,dz \\
&= \int_0^{\pi^2} \int_1^4 z^{-1/2} \sin\sqrt{yz}\,dz\,dy \\
&= \int_0^{\pi^2} \int_{\sqrt{y}}^{2\sqrt{y}} \frac{2}{\sqrt{y}} \sin u\,du\,dy \\
&= -\int_0^{\pi^2} \frac{2}{\sqrt{y}} \cos u\Big]_{u=\sqrt{y}}^{2\sqrt{y}} dy
\end{aligned}$$