About the author:
Michael Corral is an Adjunct Faculty member of the Department of Mathematics at Schoolcraft College. He received a B.A. in Mathematics from the University of California, Berkeley, and received an M.A. in Mathematics and an M.S. in Industrial & Operations Engineering from the University of Michigan.

This text was typeset in \LaTeX with the KOMA-Script bundle, using the GNU Emacs text editor on a Fedora Linux system. The graphics were created using TikZ and Gnuplot.

Copyright © 2020 Michael Corral.
Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
# Contents

## 1 The Derivative
1.1 Introduction .............................................. 1
1.2 The Derivative: Limit Approach ........................ 8
1.3 The Derivative: Infinitesimal Approach ............... 15
1.4 Derivatives of Sums, Products and Quotients ....... 21
1.5 The Chain Rule ........................................... 27
1.6 Higher Order Derivatives ............................... 32

## 2 Derivatives of Common Functions
2.1 Inverse Functions ......................................... 37
2.2 Trigonometric Functions and Their Inverses .......... 41
2.3 The Exponential and Natural Logarithm Functions .... 45
2.4 General Exponential and Logarithmic Functions .... 53

## 3 Topics in Differential Calculus
3.1 Tangent Lines ............................................. 56
3.2 Limits: Formal Definition .............................. 62
3.3 Continuity ................................................... 73
3.4 Implicit Differentiation ................................. 79
3.5 Related Rates .............................................. 82
3.6 Differentials ................................................ 84

## 4 Applications of Derivatives
4.1 Optimization ............................................... 90
4.2 Curve Sketching .......................................... 101
4.3 Numerical Approximation of Roots of Functions .... 108
4.4 The Mean Value Theorem ............................... 117

## 5 The Integral
5.1 The Indefinite Integral .................................. 123
5.2 The Definite Integral .................................... 131
5.3 The Fundamental Theorem of Calculus ................ 139
### The Greek Alphabet

<table>
<thead>
<tr>
<th>Letters</th>
<th>Name</th>
<th>Letters</th>
<th>Name</th>
<th>Letters</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>α</td>
<td>I</td>
<td>ι</td>
<td>P</td>
<td>ρ</td>
</tr>
<tr>
<td>B</td>
<td>β</td>
<td>K</td>
<td>κ</td>
<td>Σ</td>
<td>σ</td>
</tr>
<tr>
<td>Γ</td>
<td>γ</td>
<td>Λ</td>
<td>λ</td>
<td>T</td>
<td>τ</td>
</tr>
<tr>
<td>∆</td>
<td>δ</td>
<td>M</td>
<td>µ</td>
<td>Υ</td>
<td>υ</td>
</tr>
<tr>
<td>E</td>
<td>ε</td>
<td>N</td>
<td>ν</td>
<td>Φ</td>
<td>φ</td>
</tr>
<tr>
<td>Z</td>
<td>ζ</td>
<td>Ξ</td>
<td>ξ</td>
<td>X</td>
<td>χ</td>
</tr>
<tr>
<td>H</td>
<td>η</td>
<td>O</td>
<td>ο</td>
<td>Ψ</td>
<td>ψ</td>
</tr>
<tr>
<td>Θ</td>
<td>θ</td>
<td>Π</td>
<td>π</td>
<td>Ω</td>
<td>ω</td>
</tr>
</tbody>
</table>

### Mathematical Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>⇒</td>
<td>if...then; implies</td>
<td>(</td>
</tr>
<tr>
<td>⇔</td>
<td>if and only if; two-way implication</td>
<td>(</td>
</tr>
<tr>
<td>iff</td>
<td>if and only if; two-way implication</td>
<td>(</td>
</tr>
<tr>
<td>≠</td>
<td>does not imply</td>
<td>(</td>
</tr>
<tr>
<td>∃</td>
<td>there exists</td>
<td>(∃ \text{ a number } c &gt; 0)</td>
</tr>
<tr>
<td>∄</td>
<td>there does not exist</td>
<td>(\not ∃ \text{ x such that } x^2 &lt; 0)</td>
</tr>
<tr>
<td>∃!</td>
<td>there exists a unique</td>
<td>(∃! \text{ x such that } 2x - 1 = 3)</td>
</tr>
<tr>
<td>∀</td>
<td>for every</td>
<td>(∀x \geq 0, \sqrt{x} \text{ is a real number})</td>
</tr>
<tr>
<td>≡</td>
<td>is identically equal to</td>
<td>(f ≡ 0 \Rightarrow f(x) = 0 \text{ for all } x)</td>
</tr>
<tr>
<td>∝</td>
<td>is proportional to</td>
<td>(y \propto x^2 \Rightarrow y = kx^2 \text{ for some } k)</td>
</tr>
<tr>
<td>⊆</td>
<td>is a subset of</td>
<td>({0, 1} \subseteq {0, 1, 2})</td>
</tr>
<tr>
<td>∈</td>
<td>is an element of</td>
<td>(1 \in {1, 2, 3})</td>
</tr>
<tr>
<td>∉</td>
<td>is not an element of</td>
<td>(1 \notin {2, 3})</td>
</tr>
<tr>
<td>⋃</td>
<td>union of sets</td>
<td>({0, 1} \cup {2, 3} = {0, 1, 2, 3})</td>
</tr>
<tr>
<td>⋂</td>
<td>intersection of sets</td>
<td>({0, 1} \cap {1, 2} = {1})</td>
</tr>
<tr>
<td>∅</td>
<td>empty set</td>
<td>({0, 1} \cap {2, 3} = ∅)</td>
</tr>
<tr>
<td>∴</td>
<td>therefore</td>
<td>(∴ \text{ n must exist})</td>
</tr>
</tbody>
</table>
CHAPTER 1
The Derivative

1.1 Introduction

Calculus can be thought of as the analysis of curved shapes. Its development grew out of attempts to solve physical problems. For example, suppose that an object at rest 100 ft above the ground is dropped. Ignoring air resistance and wind, the object will fall straight down until it hits the ground (see Figure 1.1.1(a)). As will be proved later, $t$ seconds after being dropped the object will be $s = s(t) = -16t^2 + 100$ ft above the ground. The object will thus hit the ground after 2.5 seconds (when $s = 0$). While the object’s path is a straight line, the graph of its position $s$ above the ground as a function of time $t$ is curved, part of a parabola (see Figure 1.1.1(b)).

![Figure 1.1.1](image)

**Figure 1.1.1** An object dropped from 100 ft above the ground

How fast is the object moving before it hits the ground? This is where calculus comes in. The solution, presented now, will motivate much of this chapter.

---

1It is more than that, of course, but that definition puts us in good company: the first European textbook on calculus, written by the French mathematician Guillaume de l’Hôpital in 1696, was titled *Analyse des Infiniment Petits pour l’Intelligence des Lignes Courbes* (which translates as *Analysis of the Infinitely Small for Understanding Curved Lines*). That book (in French) can be obtained freely in electronic form at [http://www.archive.org](http://www.archive.org)
First, the object travels 100 ft in 2.5 seconds, so its **average speed** in that time is

\[
\frac{\text{distance traveled}}{\text{time elapsed}} = \frac{100 \text{ ft}}{2.5 \text{ seconds}} = 40 \text{ ft/s},
\]

and its **average velocity** in that time is

\[
\frac{\text{change in position}}{\text{change in time}} = \frac{\text{final position} - \text{initial position}}{\text{end time} - \text{start time}} = \frac{0 \text{ ft} - 100 \text{ ft}}{2.5 \text{ sec} - 0 \text{ sec}} = -40 \text{ ft/s}.
\]

Unlike speed, velocity takes direction into account. Thus, the object’s downward motion means it has negative velocity. Positive velocity implies upward motion.

Using the idea of average velocity over an interval of time, there is a natural way to define the object’s **instantaneous velocity** at a particular *instant* of time \( t \):

1. Find the average velocity over an interval of time.

2. Let the interval become smaller and smaller indefinitely, shrinking to a point \( t \). If the average velocity over that smaller and smaller interval approaches some value, call that value the instantaneous velocity at time \( t \).

Figure 1.1.2 below shows how to choose the interval: for any time \( t \) between 0 and 2.5, use the interval \([t, t + \Delta t]\), where \( \Delta t \) (pronounced “delta \( t \)”) is a small positive number. So \( \Delta t \) is the change in time over the interval; denote by \( \Delta s \) the change in the position \( s \) over that interval.

![Figure 1.1.2](image)

The average velocity of the object over the interval \([t, t + \Delta t]\) is \( \frac{\Delta s}{\Delta t} \), so since \( s(t) = -16t^2 + 100 \):
\[ \frac{\Delta s}{\Delta t} = \frac{s(t+\Delta t) - s(t)}{\Delta t} \]
\[ = \frac{-16(t+\Delta t)^2 + 100 - (-16t^2 + 100)}{\Delta t} \]
\[ = \frac{-16t^2 - 32t\Delta t - 16(\Delta t)^2 + 100 + 16t^2 - 100}{\Delta t} \]
\[ = \frac{-32t\Delta t - 16(\Delta t)^2}{\Delta t} = \frac{\Delta t(-32t - 16\Delta t)}{\Delta t} \]
\[ = -32t - 16\Delta t , \]

Now let the interval \([t, t+\Delta t]\) get smaller and smaller indefinitely—that is, let \(\Delta t\) get closer and closer to 0. Then the average velocity \(\frac{\Delta s}{\Delta t} = -32t - 16\Delta t\) gets closer and closer to \(-32t - 0 = -32t\).

Thus, the object has instantaneous velocity \(-32t\) at time \(t\). This calculation can be interpreted as taking the limit of \(\frac{\Delta s}{\Delta t}\) as \(\Delta t\) approaches 0, written as follows:

\[
\text{instantaneous velocity at } t = \text{limit of average velocity over } [t, t+\Delta t] \text{ as } \Delta t \text{ approaches to } 0 \\
= \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} \\
= \lim_{\Delta t \to 0} (-32t - 16\Delta t) \\
= -32t - 16(0) \\
= -32t 
\]

Notice that \(\Delta t\) is not replaced by 0 in the ratio \(\frac{\Delta s}{\Delta t}\) until after doing as much cancellation as possible. Notice also that the instantaneous velocity of the object varies with \(t\), as it should (why?). In particular, at the instant when the object hits the ground at time \(t = 2.5\) sec, the instantaneous velocity is \(-32(2.5) = -80\) ft/s.

If this makes sense so far, then you understand the crux of the idea of what a limit is and how to calculate a limit. The instantaneous velocity \(v(t) = -32t\) is called the derivative of the position function \(s(t) = -16t^2 + 100\). Calculating derivatives, analyzing their properties, and using them to solve various problems are part of differential calculus.

What does this have to do with curved shapes? Instantaneous velocity is a special case of an instantaneous rate of change of a function; in this case the instantaneous rate of change of the position (height above the ground) of the object. Similar to how the rate of change of a line is its slope, the instantaneous rate of change of a general curve represents the slope of the curve. For example, the parabola \(s(t) = -16t^2 + 100\) has slope \(-32t\) for all \(t\). Note that the slope of this curve varies (as a function of \(t\)), unlike the slope of a straight line.
Finding the area inside curved regions is another type of problem that calculus can solve. The basic idea is to use simpler regions—rectangles—whose areas are known, then use those to approximate the area inside the curved region. One such method is to draw more and more rectangles of diminishing widths inside the curved region, so that the sums of their areas approach the area of the curved region. Figure 1.1.3 shows an example with four rectangles to approximate the area under a curve $y = f(x)$ over an interval $[a, b]$ on which $f(x) \geq 0$.

![Figure 1.1.3 The area of a curved region](image)

The limit of these sums of rectangular areas is called an **integral**. The study and application of integrals are part of integral calculus. Perhaps the most remarkable result in calculus is that there is a connection between derivatives and integrals—the Fundamental Theorem of Calculus, discovered in the 17th century, independently, by the two men who invented calculus as we know it: English physicist, astronomer and mathematician Isaac Newton (1642-1727) and German mathematician and philosopher Gottfried Wilhelm von Leibniz (1646-1716).

Calculus makes extensive use of infinite sequences and series. An **infinite series** is just a sum of an infinite number of terms. For example, it will be shown later in the text that

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots,$$

where the sum on the right involves an infinite number of terms. A **power series** is a particular type of infinite series applied to functions; it can be thought of as a polynomial of infinite degree. For example, the trigonometric function $\sin x$ does not appear to be a polynomial. But it turns out that $\sin x$ has a power series representation as

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \cdots,$$

where again the sum continues infinitely, and the formula holds for all $x$ (in radians).

The idea of replacing a function by its power series played an important role throughout the development of calculus, and is a powerful technique in many applications.

All the functions in this text will be functions of a single real variable—that is, the values that the variable can take are real numbers. Below is some standard notation for commonly-used sets of numbers:

---

$^2$It will be shown later (in Chapter 5) that the rectangles do not have to be completely inside the region.
\[\mathbb{N} = \text{the set of all natural numbers, i.e. the set of nonnegative integers: } 0, 1, 2, 3, 4, \ldots\]

\[\mathbb{Z} = \text{the set of all integers: } 0, \pm 1, \pm 2, \pm 3, \pm 4, \ldots\]

\[\mathbb{Q} = \text{the set of all rational numbers } \frac{m}{n}, \text{ where } m \text{ and } n \text{ are integers, with } n \neq 0\]

\[\mathbb{R} = \text{the set of all real numbers}\]

Note that \(\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}\).

The set of real numbers consists of the rational numbers together with numbers that are not rational, called irrational numbers. For example, \(\sqrt{2}\) is irrational. That is, 2 is not the square of a rational number. In fact, if the square of a rational number \(q\) were an integer, then \(q\) itself would have to be an integer: write \(q = \frac{m}{n}\), where \(m\) and \(n\) are positive integers with no common positive integer divisors other than 1. Since \(q^2 = \frac{m^2}{n^2}\) simply duplicates the integer divisors of \(m\) and \(n\), then \(q^2\) can be an integer only if \(n = 1\), i.e. \(q\) is an integer. Clearly 2 is not the square of an integer, and thus it cannot be the square of a rational number. This argument also shows that \(\sqrt{3}, \sqrt{5}, \sqrt{6}, \sqrt{7}, \sqrt{8}, \sqrt{10}\), and so on, are irrational.\(^3\)

It turns out that there are far more irrational numbers—and hence real numbers—than rational numbers. In fact, whereas the rational numbers can be listed in a sequence (i.e. first, second, third, etc.), the set of real numbers cannot.\(^4\) For example, in the closed interval \([0, 1]\) there is no “next” real number after the number 0. Thus, some infinite sets are larger than others—\(\mathbb{R}\) is larger than \(\mathbb{Q}\). Intervals such as \([0, 1]\) or \(\mathbb{R}\) itself are examples of a continuum of objects, i.e. no gaps exist.\(^5\) A famous unsolved problem in mathematics—the Continuum Hypothesis—is whether an infinite set exists that is larger in size than \(\mathbb{Q}\) but smaller than \(\mathbb{R}\).

Infinity is an important notion in calculus. Whether it is the idea of infinitely large or infinitesimally small, calculus attempts to give the idea some mathematical meaning (typically by way of limits).\(^6\) This text will avoid philosophical questions about infinity.\(^7\)

Though several centuries old, calculus was the beginning of modern mathematics. Classical mathematics (e.g. algebra, geometry, trigonometry)—whose origins date back to the ancient Babylonians, Egyptians, and Greeks—was concerned mostly with the study of static quantities. Calculus produced a way to analyze dynamic (i.e. changing) quantities. The period from the 17\(^{th}\) through the 19\(^{th}\) century also saw revolutionary advances in physics, chemistry, biology and other sciences. The birth of calculus was one part of that qualitative leap.

\(^3\)This argument is due to the British philosopher Bertrand Russell (1872-1970). For an alternative proof that \(\sqrt{2}\) is irrational, see pp. 97-98 in GELFAND, I.M. AND A. SHEN, Algebra, Boston: Birkhäuser, 1993.


\(^6\)Not everyone agrees that calculus does this in a satisfactory way. For example, for an alternative development of basically the same material in “standard” calculus but without the use of limits—called infinitesimal analysis—see KEISLER, H.J., Elementary Calculus: An Infinitesimal Approach, Boston: Prindle, Weber & Schmidt, 1976.

Exercises

For Exercises 1-4, suppose that an object moves in a straight line such that its position \( s \) after time \( t \) is the given function \( s = s(t) \). Find the instantaneous velocity of the object at a general time \( t \geq 0 \). You should mimic the earlier example for the instantaneous velocity when \( s = -16t^2 + 100 \).

1. \( s = t^2 \) 
2. \( s = 9.8t^2 \) 
3. \( s = -16t^2 + 2t \) 
4. \( s = t^3 \)

5. By equation (1.1), \( \pi = 4 \left( 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots \right) \), where the \( n^{th} \) term in the sum inside the parentheses is \( \frac{(-1)^{n+1}}{2n-1} \) (starting at \( n = 1 \)). So the first approximation of \( \pi \) using this formula is \( \pi \approx 4 \left( 1 - \frac{1}{2} \right) = 8/3 \approx 2.66667 \). Continue like this until two consecutive approximations have 3 as the first digit before the decimal point. How many terms in the sum did this require? Be careful with rounding off in the approximations.

6. In elementary geometry you learned that the area inside a circle of radius \( r > 0 \) is \( \pi r^2 \) (that formula will be proved later in the text). So in particular, let \( C \) be a circle of radius 1. Then the area inside \( C \) is \( \pi \). That area can be approximated by Eudoxus' method of exhaustion. The idea is to inscribe regular polygons inside the circle, i.e. the vertexes of the polygons touch \( C \). Recall from geometry that a polygon is regular if its sides are of equal length. By increasing the number of sides of the polygons, the areas inside the polygons will approach the area (\( \pi \)) of \( C \). This was an early attempt at using what is now called a limit.

(a) Inscribe a square inside \( C \), as in Figure 1.1.4. Show that the area inside the square is 2. This is a poor approximation of \( \pi = 3.14159265... \), obviously.

(b) Inscribe a regular hexagon (6-sided) inside \( C \), as in Figure 1.1.5. Show that the area inside the hexagon is \( \frac{3\sqrt{3}}{2} \approx 2.59807621 \). This is a slightly better—though still poor—approximation of \( \pi \).

(c) Inscribe a regular dodecagon (12-sided) inside \( C \). Show that the area inside the dodecagon is 3. It thus takes 12 sides for the approximation to get the first digit of \( \pi \) correct.

(d) Inscribe a regular 100-sided polygon inside \( C \). Show that the area inside this polygon is approximately 3.13952598. This is getting closer to \( \pi \).

(e) Show that the general formula for the area inside a regular \( n \)-sided polygon inscribed inside \( C \) is \( \frac{n}{2} \sin \left( \frac{360}{n} \right) \). (Hint: The double-angle identity \( \sin 2\theta = 2 \sin \theta \cos \theta \) might help.)

---

8 This, by the way, is a terrible formula for calculating \( \pi \); getting just the 3.14 part requires 119 terms in the sum!
9 Originally due to another ancient Greek mathematician, Antiphon (ca. 430 B.C.)
10 The great ancient Greek mathematician, physicist and astronomer Archimedes (ca. 287-212 B.C.) used this method, together with circumscribed regular polygons, to calculate \( \pi \).
What is the flaw in the following “proof” that $\pi = 4$?

**Step 1:** Draw a square around a circle of diameter $d = 1$. The circumference of the circle is thus $\pi d = \pi$, and the perimeter of the square is 4.

**Step 2:** Remove corners from the square as shown in the picture on the right, so that four new corners touch the circle. Notice that the perimeter of the resulting polygon is still 4, since the lengths of the removed corner pieces are duplicated in the new polygon, so that the lengths of all the vertical sides add up to 2 while the lengths of all the horizontal sides add up to 2.

**Step 3:** Remove corners from the polygon in Step 2, as shown in the picture on the right, so that eight new corners touch the circle. The perimeter of the resulting polygon is again still 4.

**Step 4:** Continue this procedure indefinitely, with each successive polygon still having a perimeter of 4 and becoming increasingly indistinguishable from the circle. Since the perimeters of the polygons always equal 4 and approach the circle’s circumference ($\pi$), then $\pi$ must equal 4.

An infinite set is **countable** if its members can be put into a one-to-one correspondence with the members of $\mathbb{N}$, the set of natural numbers (0, 1, 2, 3, 4, ...). Clearly $\mathbb{N}$ is itself countable. The set $\mathbb{Z}$ of all integers is also countable, by means of the following one-to-one correspondence with $\mathbb{N}$:

<table>
<thead>
<tr>
<th>$\mathbb{N}$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{Z}$</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>-2</td>
<td>3</td>
<td>-3</td>
<td>4</td>
<td>-4</td>
<td>...</td>
</tr>
</tbody>
</table>

Show that $\mathbb{Q}$ (the set of all rational numbers) is countable. *(Hint: The above correspondence for $\mathbb{Z}$ is an infinite list in one dimension (the horizontal direction). For $\mathbb{Q}$ think two-dimensionally.)*
Chapter 1 • The Derivative

§1.2 The Derivative: Limit Approach

The following definition generalizes the example from the previous section (concerning instantaneous velocity) to a general function \( f(x) \):

The derivative of a real-valued function \( f(x) \), denoted by \( f'(x) \), is

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

(1.3)

for \( x \) in the domain of \( f \), provided that the limit exists.\(^\text{11}\)

For a general function \( f(x) \), the derivative \( f'(x) \) represents the instantaneous rate of change of \( f \) at \( x \), i.e. the rate at which \( f \) changes at the “instant” \( x \). For the limit part of the definition only the intuitive idea of how to take a limit—as in the previous section—is needed for now. Notice that the above definition makes the derivative \( f' \) itself a function of the variable \( x \). The function \( f' \) can be evaluated at specific values of \( x \), or you can write its general formula \( f'(x) \).

The (instantaneous) velocity of an object as the derivative of the object’s position as a function of time is only one physical application of derivatives. There are many other examples:

<table>
<thead>
<tr>
<th>Field</th>
<th>Function</th>
<th>Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>position</td>
<td>velocity</td>
</tr>
<tr>
<td></td>
<td>velocity</td>
<td>acceleration</td>
</tr>
<tr>
<td></td>
<td>momentum</td>
<td>force</td>
</tr>
<tr>
<td></td>
<td>work</td>
<td>power</td>
</tr>
<tr>
<td>Engineering</td>
<td>electric charge</td>
<td>magnetic flux</td>
</tr>
<tr>
<td></td>
<td></td>
<td>electric current</td>
</tr>
<tr>
<td></td>
<td>induced voltage</td>
<td></td>
</tr>
<tr>
<td>Economics</td>
<td>profit</td>
<td>marginal profit</td>
</tr>
</tbody>
</table>

The limit definition can be used for finding the derivatives of simple functions.

**Example 1.1**

Find the derivative of the function \( f(x) = 1 \).

*Solution:* By definition, \( f(x) = 1 \) for all \( x \), so:

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

\[
= \lim_{\Delta x \to 0} \frac{1 - 1}{\Delta x}
\]

\[
= \lim_{\Delta x \to 0} \frac{0}{\Delta x}
\]

\[
= \lim_{\Delta x \to 0} 0
\]

\[
f'(x) = 0
\]

\(^{11}\)Recall that the domain of \( f \) is the set of all numbers \( x \) such that \( f(x) \) is defined.
Notice in the above example that replacing $\Delta x$ by 0 was unnecessary when taking the limit, since the ratio $\frac{f(x+\Delta x) - f(x)}{\Delta x}$ simplified to 0 before taking the limit, and the limit of 0 is 0 regardless of what $\Delta x$ approaches. In fact, the answer—namely, $f'(x) = 0$ for all $x$—should have been obvious without any calculations: the function $f(x) = 1$ is a constant function, so its value (1) never changes, and thus its rate of change is always 0. Hence, its derivative is 0 everywhere. Replacing the constant 1 by any constant yields the following important result:

The derivative of any constant function is 0.

The above discussion shows that the calculation in Example 1.1 was unnecessary. Consider another example where no calculation is required to find the derivative: the function $f(x) = x$. The graph of this function is just the line $y = x$ in the $xy$-plane, and the rate of change of a line is a constant, called its slope. The line $y = x$ has a slope of 1, so the derivative of $f(x) = x$ is $f'(x) = 1$ for all $x$. The formal calculation of the derivative, though unnecessary, verifies this:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{(x+\Delta x) - x}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x} = \lim_{\Delta x \to 0} 1 = 1$$

Recall that a function whose graph is a line is called a linear function. For a general linear function $f(x) = mx + b$, where $m$ is the slope of the line and $b$ is its $y$-intercept, the same argument as above for $f(x) = x$ yields the following result:

The derivative of any linear function is the slope of the line itself:
If $f(x) = mx + b$ then $f'(x) = m$ for all $x$.

The function $f(x) = 1$ from Example 1.1 is the special case where $m = 0$ and $b = 1$; its graph is a horizontal line, so its slope (and hence its derivative) is 0 for all $x$. Likewise, the function $f(x) = 2x - 1$ represents a line of slope $m = 2$, so its derivative is 2 for all $x$. Figure 1.2.1 shows these and other linear functions $y = f(x)$.

![Figure 1.2.1](image-url)  
**Figure 1.2.1** Slopes and derivatives of lines
Linear functions have a constant derivative—the constant being the slope of the line. The converse turns out to be true: a function with a constant derivative must be a linear function.\(^{12}\)

What types of functions do not have constant derivatives? The previous section discussed such a function: the parabola \(s(t) = -16t^2 + 100\), whose derivative \(s'(t) = -32t\) is clearly not a constant function. In general, functions that represent curves (i.e. not straight lines) do not change at a constant rate—that is precisely what makes them curved. So such functions do not have a constant derivative.

**Example 1.2**

Find the derivative of the function \(f(x) = \frac{1}{x}\). Also, find the instantaneous rate of change of \(f\) at \(x = 2\).

**Solution:** For all \(x \neq 0\), the derivative \(f'(x)\) is:

\[
\begin{align*}
  f'(x) &= \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \\
  &= \lim_{\Delta x \to 0} \frac{1}{x + \Delta x} - \frac{1}{x} \\
  &= \lim_{\Delta x \to 0} \frac{x - (x + \Delta x)}{\Delta x(x + \Delta x)} \\
  &= \lim_{\Delta x \to 0} \frac{-\Delta x}{\Delta x(x + \Delta x)} \\
  &= \lim_{\Delta x \to 0} \frac{-1}{(x + \Delta x)x} = \frac{-1}{(x + 0)x} \\
  f'(x) &= -\frac{1}{x^2}
\end{align*}
\]

The instantaneous rate of change of \(f\) at \(x = 2\) is just the derivative \(f'(x)\) evaluated at \(x = 2\), that is, \(f'(2) = -\frac{1}{4}\).

Notice that the instantaneous rate of change \(f'(2) = -\frac{1}{4}\) in the above example is a negative number. This should make sense, since the function \(f(x) = \frac{1}{x}\) is changing in the negative direction at \(x = 2\); that is, \(f(x)\) is decreasing in value at \(x = 2\). This is plain to see from the graph of \(f(x) = \frac{1}{x}\) shown on the right. In fact, for all \(x \neq 0\) the function \(f(x) = \frac{1}{x}\) is decreasing as \(x\) grows.\(^{13}\) This is reflected in the derivative \(f'(x) = -\frac{1}{x^2}\) being negative for all \(x \neq 0\). In general, a negative derivative means that the function is decreasing, while a positive derivative means that it is increasing.

\(^{12}\)This will be proved in Chapter 5.

\(^{13}\)In this text, the rate of change of \(f(x)\) is always taken in the direction of increasing \(x\), i.e. in the positive \(x\) direction.
The problem with using the limit definition to find the derivative of a curved function is that the calculations require more work, as the above example shows. As the functions become more complicated those calculations can become difficult or even impossible. And though limits have not yet been defined formally, for now the intuitively obvious idea of limits suffices, namely:

For a real number \( a \) and a real-valued function \( f(x) \), say that the limit of \( f(x) \) as \( x \) approaches \( a \) equals the number \( L \), written as

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

if \( f(x) \) approaches \( L \) as \( x \) approaches \( a \).

Equivalently, this means that \( f(x) \) can be made as close as you want to \( L \) by choosing \( x \) close enough to \( a \). Note that \( x \) can approach \( a \) from any direction.

Below are some simple rules for limits, which will be proved later:

**Rules for Limits:** Suppose that \( a \) is a real number and that \( f(x) \) and \( g(x) \) are real-valued functions such that \( \lim_{x \to a} f(x) \) and \( \lim_{x \to a} g(x) \) both exist. Then:

(a) \[ \lim_{x \to a} (f(x) + g(x)) = \left( \lim_{x \to a} f(x) \right) + \left( \lim_{x \to a} g(x) \right) \]

(b) \[ \lim_{x \to a} (f(x) - g(x)) = \left( \lim_{x \to a} f(x) \right) - \left( \lim_{x \to a} g(x) \right) \]

(c) \[ \lim_{x \to a} (k \cdot f(x)) = k \cdot \left( \lim_{x \to a} f(x) \right) \text{ for any constant } k \]

(d) \[ \lim_{x \to a} (f(x) \cdot g(x)) = \left( \lim_{x \to a} f(x) \right) \cdot \left( \lim_{x \to a} g(x) \right) \]

(e) \[ \lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)} \text{, if } \lim_{x \to a} g(x) \neq 0 \]

The above rules say that the limit of sums, differences, constant multiples, products, and quotients is the sum, difference, constant multiple, product, and quotient, respectively, of the limits. This seems intuitively obvious.

These rules can be used for finding other expressions for the derivative. The quantity \( \Delta x \) represents a small number—positive or negative—that approaches 0, but it is common in mathematics texts to use the letter \( h \) instead:

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

\[14\] Physics texts typically prefer the delta notation, since \( \Delta x \) represents a small change in some physical quantity \( x \).
Another formulation is to set \( h = w - x \) in formula (1.4), which yields

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

so that

\[
f'(x) = \lim_{w - x \to 0} \frac{f(w) - f(x)}{w - x},
\]

(1.5)

since \( w - x \) approaches 0 if and only if \( w \) approaches \( x \). Another formulation replaces \( h \) by \( -h \):

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

and thus

\[
f'(x) = \lim_{h \to 0} \frac{f(x) - f(x - h)}{-2h},
\]

(1.6)

since \( -h \) approaches 0 if and only if \( h \) approaches 0. The above formulations did not use the Limit Rules, but the following result does:

Suppose that \( f'(x) \) exists. Then

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

(1.7)

Proof: Since \( f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{f(x) - f(x - h)}{h} \) by formulas (1.4) and (1.6), then Limit Rule (c) shows that

\[
\frac{1}{2} f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{2h} = \lim_{h \to 0} \frac{f(x) - f(x - h)}{2h}.
\]

Now use the idea that \( a - b = (a - c) + (c - b) \) for all \( a, b, \) and \( c \) to write:

\[
\lim_{h \to 0} \frac{f(x + h) - f(x - h)}{2h} = \lim_{h \to 0} \frac{(f(x + h) - f(x)) + (f(x) - f(x - h))}{2h}
\]

\[
= \lim_{h \to 0} \frac{f(x + h) - f(x)}{2h} + \lim_{h \to 0} \frac{f(x) - f(x - h)}{2h} \quad \text{(by Limit Rule (a))}
\]

\[
= \frac{1}{2} f'(x) + \frac{1}{2} f'(x)
\]

\[
= f'(x)
\]

QED
As an example of using these different formulations, recall that a function $f$ is **even** if $f(-x) = f(x)$ for all $x$ in the domain of $f$, and $f$ is **odd** if $f(-x) = -f(x)$ for all $x$ in its domain. For example, $x^2$, $x^4$, and $\cos x$ are even functions; $x$, $x^3$, and $\sin x$ are odd functions. The following result is often useful:

![The derivative of an even function is an odd function.
The derivative of an odd function is an even function.]

To prove the first statement—the second is an exercise—suppose that $f$ is an even function and that $f'(x)$ exists for all $x$ in its domain. Then

$$f'(-x) = \lim_{h \to 0} \frac{f(-x + h) - f(-x)}{h} \quad \text{by formula (1.4) with } x \text{ replaced by } -x$$

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

$$= \lim_{h \to 0} \frac{f(x-h) - f(x)}{h} \quad \text{since } f \text{ is even}$$

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

$$= -\lim_{h \to 0} \frac{f(x) - f(x-h)}{h} \quad \text{by Limit Rule (c), so}$$

$$f'(-x) = -f'(x) \quad \text{by formula (1.6)},$$

which shows that $f'$ is an odd function.

Derivatives do not always exist, as the following example shows.

**Example 1.3**

Let $f(x) = |x|$. Show that $f'(0)$ does not exist.

*Solution:* Recall that the absolute value function $f(x) = |x|$ is defined as

$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

The graph consists of two lines meeting at the origin. For $x \geq 0$ the graph is the line $y = x$, which has slope 1. For $x \leq 0$ the graph is the line $y = -x$, which has slope -1. These lines agree in value ($y = 0$) at $x = 0$, but their slopes do not agree in value at $x = 0$. Therefore the derivative of $f$ does not exist at $x = 0$, since the derivative of a curve is just its slope. A more “formal” proof (which amounts to the same argument) is outlined in the exercises.
If the derivative \( f'(x) \) exists then \( f \) is \textbf{differentiable} at \( x \). A \textbf{differentiable function} is one that is differentiable at every point in its domain. For example, \( f(x) = x \) is a differentiable function, but \( f(x) = |x| \) is not differentiable at \( x = 0 \). The act of calculating a derivative is called \textbf{differentiation}. For example, differentiating the function \( f(x) = x \) yields \( f'(x) = 1 \).

### Exercises

Note: For all exercises, you can use anything discussed so far (including previous exercises).

**A**

For Exercises 1-11, find the derivative of the given function \( f(x) \) for all \( x \) (unless indicated otherwise).

1. \( f(x) = 0 \)
2. \( f(x) = 1 - 3x \)
3. \( f(x) = (x + 1)^2 \)
4. \( f(x) = 2x^2 - 3x + 1 \)
5. \( f(x) = 
\frac{1}{x+1}, \text{ for all } x \neq -1 \)
6. \( f(x) = 
\frac{1}{x+1}, \text{ for all } x \neq -1 \)
7. \( f(x) = 
\frac{1}{x+1}, \text{ for all } x \neq 0 \)
8. \( f(x) = \sqrt{x}, \text{ for all } x > 0 \) (Hint: Rationalize the numerator in the definition of the derivative.)
9. \( f(x) = \sqrt{x+1}, \text{ for all } x > -1 \)
10. \( f(x) = \sqrt{x^2+1} \)
11. \( f(x) = \sqrt{x^2+3x+4} \)

12. In Exercise 8 the point \( x = 0 \) was excluded when calculating \( f'(x) \), even though \( x = 0 \) is in the domain of \( f(x) = \sqrt{x} \). Can you explain why \( x = 0 \) was excluded?

**B**

13. Show that for all functions \( f \) such that \( f'(x) \) exists, \( f'(x) = \lim_{w \to x} \frac{f(x) - f(w)}{x - w} \).
14. True or false: If \( f \) and \( g \) are differentiable functions on an interval \( (a, b) \) and \( f(x) < g(x) \) for all \( x \) in \( (a, b) \), then \( f'(x) < g'(x) \) for all \( x \) in \( (a, b) \). If true, prove it; if false, give a counterexample. Would your answer change if the restriction of \( x \) to \( (a, b) \) were removed and all real \( x \) were used instead?
15. Show that the derivative of an odd function is an even function.

**C**

For Exercises 16-21, assuming that \( f'(x) \) exists, prove the given formula.

16. \( f'(x) = \lim_{h \to 0} \frac{f(x+2h) - f(x-2h)}{4h} \)
17. \( f'(x) = \lim_{h \to 0} \frac{f(x+3h) - f(x-3h)}{6h} \)
18. \( f'(x) = \lim_{h \to 0} \frac{f(x+2h) - f(x-3h)}{5h} \)
19. \( f'(x) = \lim_{h \to 0} \frac{f(x+ah) - f(x-bh)}{(a+b)h} \) \( (a, b > 0) \)

20. \( \lim_{w \to x} \frac{w f(x) - x f(w)}{w - x} = f(x) - x f'(x) \)
21. \( \lim_{w \to x} \frac{w^2 f(x) - x^2 f(w)}{w - x} = 2xf(x) - x^2 f'(x) \)

22. Show that \( f(x) = |x| \) is not differentiable at \( x = 0 \), using formula (1.4) for the derivative. Here you will have to use a part of the definition which has not been used yet: as \( h \) approaches 0, \( h \) can be either positive or negative. Consider those two cases in showing that the limit is not defined at \( x = 0 \).
23. Suppose that \( f(a + b) = f(a)f(b) \) for all \( a \) and \( b \), and \( f'(0) \) exists. Show that \( f'(x) \) exists for all \( x \).
1.3 The Derivative: Infinitesimal Approach

Traditionally a function \( f \) of a variable \( x \) is written as \( y = f(x) \). The dependent variable \( y \) is considered a function of the independent variable \( x \). This allows taking the derivative of \( y \) with respect to \( x \), i.e. the derivative of \( y \) as a function of \( x \), denoted by \( \frac{dy}{dx} \). This is simply a different way of writing \( f'(x) \), and is just one of many ways of denoting the derivative:

**Notation for the derivative of \( y = f(x) \):** The following are all equivalent:

\[
\frac{dy}{dx}, \quad f'(x), \quad \frac{d}{dx} (f(x)), \quad y', \quad \dot{y}, \quad \dot{f}(x), \quad \frac{df}{dx}, \quad Df(x)
\]

The notation \( \frac{dy}{dx} \) appears to denote a fraction: a quantity \( dy \) divided by a quantity \( dx \). It turns out that the derivative really can be thought of in that way, as a ratio of infinitesimals. In fact, this was the way in which derivatives were used by the founders of calculus—Newton and, in particular, Leibniz.\(^{15}\) Even today, this is often the way in which derivatives are thought of and used in fields outside of mathematics, such as physics, engineering, and chemistry, perhaps due to its more intuitive nature.

The concept of infinitesimals used here is based on the nilsquare infinitesimal approach developed by J.L. Bell\(^ {16}\), namely:

A number \( \delta \) is an **infinitesimal** if the conditions (a)-(d) hold:

(a) \( \delta \neq 0 \)

(b) if \( \delta > 0 \) then \( \delta \) is smaller than any positive real number

(c) if \( \delta < 0 \) then \( \delta \) is larger than any negative real number

(d) \( \delta^2 = 0 \) (and hence all higher powers of \( \delta \), such as \( \delta^3 \) and \( \delta^4 \), are also 0)

Note: Any infinitesimal multiplied by a nonzero real number is also an infinitesimal, while 0 times an infinitesimal is 0.

The above definition says that infinitesimals are numbers which are closer to 0 than any positive or negative number without being zero themselves, and raising them to powers greater than or equal to 2 makes them 0. So infinitesimals are not real numbers.\(^ {17}\) This is not a problem, since calculus deals with other numbers, such as infinity, which are not real. An infinitesimal can be thought of as an infinitely small number arbitrarily close to 0 but not 0.

\(^ {15}\)It was Leibniz who created the notation \( \frac{dy}{dx} \). For this reason \( \frac{dy}{dx} \) is called the Leibniz notation for the derivative. Newton used the dot notation \( \dot{y} \), which has fallen out of favor with mathematicians but is still used by many physicists, especially when the independent variable represents time. Newton called derivatives fluxions. The prime notation \( f' \) is due to the French mathematician and physicist Joseph Louis Lagrange (1736-1813).


\(^ {17}\)In an equivalent treatment, infinitesimals are part of the hyperreal number system. See KEISLER, H.J., *Elementary Calculus: An Infinitesimal Approach*, Boston: Prindle, Weber & Schmidt, 1976.
This might seem like a strange notion, but it really is not all that different from the limit notion where, say, you let \( \Delta x \) approach 0 but not necessarily let it equal 0.\(^{18}\)

As for the square of a nonzero infinitesimal being 0, think of how a calculator handles the squares of small numbers. For example, most calculators can display \( 10^{-8} \) as 0.00000001, and will even let you add 1 to that to get 1.00000001. But when you square \( 10^{-8} \) and add 1 to it, most calculators will display the sum as simply 1. The calculator treats the square of \( 10^{-8} \), namely \( 10^{-16} \), as a number so small compared to 1 that it is effectively zero.\(^{19}\)

Notice a major difference between 0 and an infinitesimal \( \delta \): \( 2 \cdot 0 \) and 0 are the same, but \( 2 \delta \) and \( \delta \) are distinct. This holds for any nonzero constant multiple, not just the number 2.

The derivative \( \frac{dy}{dx} \) of a function \( y = f(x) \) can now be defined in terms of infinitesimals:

Let \( dx \) be an infinitesimal such that \( f(x + dx) \) is defined. Then \( dy = f(x + dx) - f(x) \) is also an infinitesimal, and the derivative of \( y = f(x) \) at \( x \) is the ratio of \( dy \) to \( dx \):

\[
\frac{dy}{dx} = \frac{f(x + dx) - f(x)}{dx} \quad (1.8)
\]

The basic idea is that \( dx \) is an infinitesimally small change in the variable \( x \), producing an infinitesimally small change \( dy \) in the value of \( y = f(x) \).

**Example 1.4**

Show that the derivative of \( y = f(x) = x^2 \) is \( \frac{dy}{dx} = 2x \).

*Solution:* For any real number \( x \),

\[
\frac{dy}{dx} = \frac{f(x + dx) - f(x)}{dx} = \frac{(x + dx)^2 - x^2}{dx} = \frac{x^2 + 2x dx + (dx)^2 - x^2}{dx} = \frac{2x dx + 0}{dx} \quad \text{since } dx \text{ is an infinitesimal } \Rightarrow (dx)^2 = 0
\]

\[
= \frac{2x dx}{dx} = 2x
\]

---

\(^{18}\)The infinitesimal approach was first developed in an axiomatic manner in the landmark book ROBINSON, A., *Non-Standard Analysis*, Amsterdam: North-Holland, 1966. Robinson showed that for all practical purposes calculus can be developed without resorting to limits, with equivalent results.

\(^{19}\)Calculators do this for display reasons—most can show only 10-12 digits. Try this experiment on your calculator: Add \( 10^{30} \), \(-10^{30}\), and 1 in two different ways: \((10^{30} + (-10^{30})) + 1\), and \(10^{30} + (-10^{30}) + 1\). The first way will give you the correct answer 1, but the second way yields 0. So addition is not always associative on calculators!
You might have noticed that the above example did not involve limits, and that the derivative \(2x\) represents a real number (i.e. no infinitesimals appear in the final answer); this will always be the case. Infinitesimals possess another useful property:

**Microstraightness Property:** For the graph of a differentiable function, any part of the curve with infinitesimal length is a straight line segment.

In other words, *at the infinitesimal level differentiable curves are straight.* The idea behind this is simple. At various points on a nonstraight differentiable curve \(y = f(x)\) the distances along the curve between the points are not quite the same as the lengths of the line segments joining the points. For example, in Figure 1.3.1 the distance \(s\) measured along the curve from the point \(A\) to the point \(B\) is not the same as the length of the line segment \(AB\) joining \(A\) to \(B\).

![Figure 1.3.1](image)

**Figure 1.3.1** Curved real-valued distance \(s \neq\) length of \(AB\), but infinitesimal \(s =\) length of \(AB\)

However, as the points \(A\) and \(B\) get closer to each other, the difference between that part of the curve joining \(A\) to \(B\) and the line segment \(AB\) becomes less noticeable. That is, the curve is *almost* linear when \(A\) and \(B\) are close. The Microstraightness Property simply goes one step further and says that the curve actually *is* linear when the distance \(s\) between the points is infinitesimal (so that \(s\) equals the length of \(AB\) at the infinitesimal level).

At first this might seem nonsensical. After all, how could any nonstraight part of a curve be straight? You have to remember that an infinitesimal is an abstraction—it does not exist physically. A curve \(y = f(x)\) is also an abstraction, which exists in a purely mathematical sense, so its geometric properties at the “normal” scale do not have to match those at the infinitesimal scale (which can be defined in any way, provided the properties at that scale are consistent).

This abstraction finally reveals what an instantaneous rate of change is: the average rate of change over an infinitesimal interval. Moving an infinitesimal amount \(dx\) away from a value \(x\) produces an infinitesimal change \(dy\) in a differentiable function \(y = f(x)\). The average rate of change of \(y = f(x)\) over the infinitesimal interval \([x, x + dx]\) is thus \(\frac{dy}{dx}\), i.e. the slope—rise over run—of the straight line segment represented by the curve \(y = f(x)\) over that interval, as in the figure on the right.\(^{20}\)

\(^{20}\) Notice that the figure implies that the Pythagorean Theorem does not apply to infinitesimal triangles. This will be discussed in Chapter 7.
The Microstraightness Property can be extended to smooth curves—that is, curves without sharp edges or cusps. For example, circles and ellipses are smooth, but polygons are not.

The properties of infinitesimals can be applied to determine the derivatives of the sine and cosine functions. Consider a circle of radius 1 with center \( O \) and points \( A \) and \( B \) on the circle such that the line segment \( AB \) is a diameter. Let \( C \) be a point on the circle such that the angle \( \angle BAC \) has an infinitesimal measure \( dx \) (in radians) as in Figure 1.3.2(a).

\[
\text{Figure 1.3.2} \quad \text{Circle } O: BC = 2 \sin dx, \ \angle BOC = 2 \angle BAC
\]

By Thales’ Theorem from elementary geometry, the angle \( \angle ACB \) is a right angle. Thus:

\[
\sin dx = \frac{BC}{AB} = \frac{BC}{2} \implies BC = 2 \sin dx
\]

Figure 1.3.2(b) shows that \( \angle OAC + \angle OCA + \angle AOC = \pi \). Thus, \( 1 = OC = OA \implies \angle OCA = \angle OAC = dx \implies \angle OAC = \pi - dx - dx = \pi - 2dx \implies \angle BOC = 2dx \). By the arc length formula from trigonometry, the length \( s \) of the arc \( BC \) along the circle from \( B \) to \( C \) is the radius times the central angle \( \angle BOC \):

\[
s = BC = 2dx = 2 \sin dx
\]

Since \( dx \) is an infinitesimal, \( (dx)^2 = 0 \). So since \( \sin^2 dx + \cos^2 dx = 1 \), then:

\[
\cos^2 dx = 1 - \sin^2 dx = 1 - (dx)^2 = 1 - 0 = 1 \implies \cos dx = 1
\]

The derivative of \( y = \sin x \) is then:

\[
\frac{d}{dx} (\sin x) = \frac{dy}{dx} = \frac{\sin(x + dx) - \sin x}{dx} = \frac{(\sin x \cos dx + \sin dx \cos x) - \sin x}{dx} \quad \text{by the sine addition formula}
\]

\[
= \frac{(\sin x)(1) + dx \cos x - \sin x}{dx} = \frac{dx \cos x}{dx} \quad \text{by parts}
\]

and thus:

\[
\frac{d}{dx} (\sin x) = \cos x
\]
A similar argument (left as an exercise) using the cosine addition formula shows:

\[
\frac{d}{dx} (\cos x) = -\sin x
\]

One of the intermediate results proved here bears closer examination. Namely, \( \sin dx = dx \) for an infinitesimal angle \( dx \) measured in radians. At first, it might seem that this cannot be true. After all, an infinitesimal \( dx \) is thought of as being infinitely close to 0, and \( \sin 0 = 0 \), so you might expect that \( \sin dx = 0 \). But this is not the case. The formula \( \sin dx = dx \) says that in an infinitesimal interval around 0, the function \( y = \sin x \) is identical to the line \( y = x \) (not the line \( y = 0 \)). This, in turn, suggests that for real-valued \( x \) close to 0, \( \sin x \approx x \).

This indeed turns out to be the case. The free graphing software Gnuplot\(^{21}\) can display the graphs of \( y = \sin x \) and \( y = x \). Figure 1.3.3(a) below shows how those graphs compare over the interval \([-\pi, \pi]\). Outside the interval \([-1, 1]\) there is a noticeable difference.

Figure 1.3.3(b) shows that there is virtually no difference in the graphs even over the non-infinitesimal interval \([-0.3, 0.3]\). So \( \sin x \approx x \) is a very good approximation when \( x \) is close to 0, that is, when \( |x| \ll 1 \) (the symbol \( \ll \) means “much less than”). This approximation is used in many applications in engineering and physics when the angle \( x \) is assumed to be small.

Notice something else suggested by the relation \( \sin dx = dx \): there is a fundamental difference at the infinitesimal level between a line of slope 1 (\( y = x \)) and a line of slope 0 (\( y = 0 \)). In a real interval \((-a, a)\) around \( x = 0 \) the difference between the two lines can be made as small as desired by choosing \( a > 0 \) small enough. But in an infinitesimal interval \((-\delta, \delta)\) around \( x = 0 \) there is unbridgeable gulf between the two lines. This is the crucial difference in \( \sin dx \) being equal to \( dx \) rather than 0.

\(^{21}\) Available at http://www.gnuplot.info.
Notice also that the value of a function at an infinitesimal may itself be an infinitesimal (e.g. \( \sin dx = dx \)) or a real number (e.g. \( \cos dx = 1 \)).

For a differentiable function \( f(x) \), \( \frac{df}{dx} = f'(x) \) and so multiplying both sides by \( dx \) yields the important relation:

\[
df = f'(x) \, dx
\]  

(1.9)

Note that both sides of the above equation are infinitesimals for each value of \( x \) in the domain of \( f' \), since \( f'(x) \) would then be a real number.

The notion of an infinitesimal was fairly radical at the time (and still is). Some mathematicians embraced it, e.g. the outstanding Swiss mathematician Leonhard Euler (1707-1783), who produced a large amount of work using infinitesimals. But it was too radical for many mathematicians (and philosophers\(^{22} \)), enough so that by the 19\(^{th} \) century some mathematicians (notably Augustin Cauchy and Karl Weierstrass) felt the need to put calculus on what they considered a more “rigorous” footing, based on limits.\(^{23} \) Yet it was precisely the notion of an infinitesimal which lent calculus its modern character, by showing the power and usefulness of such an abstraction (especially one that did not obey the rules of classical mathematics).

### Exercises

**A**

For Exercises 1-9, let \( dx \) be an infinitesimal and prove the given formula.

1. \( (dx + 1)^2 = 2dx + 1 \)
2. \( (dx + 1)^3 = 3dx + 1 \)
3. \( (dx + 1)^4 = 4dx + 1 \)
4. \( \tan dx = dx \)
5. \( \sin 2dx = 2dx \)
6. \( \cos 2dx = 1 \)
7. \( \sin 3dx = 3dx \)
8. \( \cos 3dx = 1 \)
9. \( \sin 4dx = 4dx \)

10. Is \( \cot dx \) defined for an infinitesimal \( dx \)? If so, then find its value. If not, then explain why.

11. In the proof of the derivative formulas for \( \sin x \) and \( \cos x \), the equation \( \cos^2 dx = 1 \) was solved to give \( \cos dx = 1 \). Why was the other possible solution \( \cos dx = -1 \) ignored?

**B**

12. Show that \( \frac{d}{dx} (\cos x) = -\sin x \).

13. Show that \( \frac{d}{dx} (\cos 2x) = -2 \sin 2x \). (Hint: Use Exercises 5 and 6.)

**C**

14. Show that \( \frac{d}{dx} (\tan x) = \sec^2 x \). (Hint: Use Exercise 4.)

\(^{22}\)The English philosopher George Berkeley (1685-1753) famously derided infinitesimals as “the ghosts of departed quantities” in his book *The Analyst* (1734), which had the disquieting subtitle “A Discourse Addressed to an Infidel Mathematician” (directed at Newton).

\(^{23}\)However, the limit approach turns out, ultimately, to be equivalent to the infinitesimal approach. In essence, only the terminology is different.
1.4 Derivatives of Sums, Products and Quotients

So far the derivatives of only a few simple functions have been calculated. The following rules will make it easier to calculate derivatives of more functions:

**Rules for Derivatives:** Suppose that \( f \) and \( g \) are differentiable functions of \( x \). Then:

- **Sum Rule:** \( \frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx} \)
- **Difference Rule:** \( \frac{d}{dx}(f - g) = \frac{df}{dx} - \frac{dg}{dx} \)
- **Constant Multiple Rule:** \( \frac{d}{dx}(cf) = c \cdot \frac{df}{dx} \) for any constant \( c \)
- **Product Rule:** \( \frac{d}{dx}(f \cdot g) = f \cdot \frac{dg}{dx} + g \cdot \frac{df}{dx} \)
- **Quotient Rule:** \( \frac{d}{dx} \left( \frac{f}{g} \right) = \frac{g \cdot \frac{df}{dx} - f \cdot \frac{dg}{dx}}{g^2} \)

The above rules can be written using the prime notation for derivatives:

- **Sum Rule:** \((f + g)'(x) = f'(x) + g'(x)\)
- **Difference Rule:** \((f - g)'(x) = f'(x) - g'(x)\)
- **Constant Multiple Rule:** \((cf)'(x) = c \cdot f'(x)\) for any constant \( c \)
- **Product Rule:** \((f \cdot g)'(x) = f(x) \cdot g'(x) + g(x) \cdot f'(x)\)
- **Quotient Rule:** \( \left( \frac{f}{g} \right)'(x) = \frac{g(x) \cdot f'(x) - f(x) \cdot g'(x)}{(g(x))^2} \)

The proof of the Sum Rule is straightforward. Since \( \frac{df}{dx} \) and \( \frac{dg}{dx} \) both exist then:

\[
\frac{d}{dx} (f + g) = \frac{(f + g)(x + dx) - (f + g)(x)}{dx} = \frac{f(x + dx) + g(x + dx) - (f(x) + g(x))}{dx} \\
= \frac{f(x + dx) - f(x) + g(x + dx) - g(x)}{dx} = \frac{f(x + dx) - f(x)}{dx} + \frac{g(x + dx) - g(x)}{dx} \\
= \frac{df}{dx} + \frac{dg}{dx} \quad \checkmark
\]

The proofs of the Difference and Constant Multiple Rules are similar and are left as exercises.
Note that by the Product Rule, in general the derivative of a product is not the product of the derivatives. That is, \( \frac{d(f \cdot g)}{dx} \neq \frac{df}{dx} \cdot \frac{dg}{dx} \). This should be obvious from some earlier examples. For instance, if \( f(x) = x \) and \( g(x) = 1 \) then \( (f \cdot g)(x) = x \cdot 1 = x \) so that \( \frac{d(f \cdot g)}{dx} = 1 \), but \( \frac{df}{dx} \cdot \frac{dg}{dx} = 1 \cdot 0 = 0 \).

There is a proof of the Product Rule similar to the proof of the Sum Rule (see Exercise 20), but there is a more geometric way of seeing why the formula holds, described below.

Construct a rectangle whose perpendicular sides have lengths \( f(x) \) and \( g(x) \) for some \( x \), as in the drawing on the right. Change \( x \) by some infinitesimal amount \( dx \), which produces infinitesimal changes \( df \) and \( dg \) in \( f(x) \) and \( g(x) \), respectively. Assume those changes are positive and extend the original rectangle by those amounts, creating a larger rectangle with perpendicular sides \( f(x+dx) \) and \( g(x+dx) \). Then

\[
d(f \cdot g) = (f \cdot g)(x + dx) - (f \cdot g)(x)
= f(x + dx) \cdot g(x + dx) - f(x) \cdot g(x)
= \text{(area of outer rectangle)} - \text{(area of original rectangle)}
= \text{sum of the areas of the three shaded inner rectangles}
= f(x) \cdot dg + g(x) \cdot df + df \cdot dg
= f(x) \cdot dg + g(x) \cdot df + (f'(x) dx) \cdot (g'(x) dx)
= f(x) \cdot dg + g(x) \cdot df + (f'(x) g'(x)) \cdot (dx)^2
= f(x) \cdot dg + g(x) \cdot df + (f'(x) g'(x)) \cdot 0
\]

\[
d(f \cdot g) = f(x) \cdot dg + g(x) \cdot df
\]

so dividing both sides by \( dx \) yields

\[
\frac{d(f \cdot g)}{dx} = f(x) \cdot \frac{dg}{dx} + g(x) \cdot \frac{df}{dx} \quad \checkmark
\]

To prove the Quotient Rule, let \( y = \frac{f}{g} \), so \( f = g \cdot y \). If \( y \) were a differentiable function of \( x \), then the Product Rule would give

\[
\frac{df}{dx} = \frac{d(g \cdot y)}{dx} = g \cdot \frac{dy}{dx} + y \cdot \frac{dg}{dx} = g \cdot \frac{dy}{dx} + \frac{f}{g} \cdot \frac{dg}{dx} \Rightarrow g \cdot \frac{dy}{dx} = \frac{df}{dx} - \frac{f}{g} \cdot \frac{dg}{dx}
\]

and so dividing both sides by \( g \) and getting a common denominator gives

\[
\frac{dy}{dx} = \frac{1}{g} \cdot \frac{df}{dx} - \frac{f}{g^2} \cdot \frac{dg}{dx} = g \cdot \frac{df}{dx} - f \cdot \frac{dg}{dx} \quad \checkmark
\]

A simple mnemonic device for remembering the Quotient Rule is: write \( \frac{f}{g} \) as \( \frac{HI}{HO} \)—so that HI represents the “high” (numerator) part of the quotient and HO represents the “low” (denominator) part—and think of \( dHI \) and \( dHO \) as the derivatives of HI and HO, respectively. Then

\[
\frac{d}{dx} \left( \frac{f}{g} \right) = \frac{HO \cdot dHI - HI \cdot dHO}{HO^2}, \text{ pronounced as “ho-dee-hi minus hi-dee-ho over ho-ho.”}
\]
Example 1.5

Use the Quotient Rule to show that \( \frac{d}{dx}(\tan x) = \sec^2 x \).

**Solution:** Since \( \tan x = \frac{\sin x}{\cos x} \) then:

\[
\frac{d}{dx}(\tan x) = \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) = \frac{(\cos x) \cdot \frac{d}{dx}(\sin x) - (\sin x) \cdot \frac{d}{dx}(\cos x)}{\cos^2 x} = \frac{(\cos x)(\cos x) - (\sin x)(-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x
\]

Example 1.6

Use the Quotient Rule to show that \( \frac{d}{dx}(\sec x) = \sec x \tan x \).

**Solution:** Since \( \sec x = \frac{1}{\cos x} \) then:

\[
\frac{d}{dx}(\sec x) = \frac{d}{dx}\left(\frac{1}{\cos x}\right) = \frac{\cos x \cdot \frac{d}{dx}(1) - 1 \cdot \frac{d}{dx}(\cos x)}{\cos^2 x} = \frac{(\cos x)(0) - 1 \cdot (-\sin x)}{\cos^2 x} = \frac{\sin x}{\cos^2 x} = \frac{1}{\cos x} \cdot \frac{\sin x}{\cos x} = \sec x \tan x
\]

Similar to the above examples, the derivatives of \( \cot x \) and \( \csc x \) can be found using the Quotient Rule (left as exercises). The derivatives of all six trigonometric functions are:

\[
\begin{align*}
\frac{d}{dx}(\cos x) &= -\sin x & \frac{d}{dx}(\sec x) &= \sec x \tan x \\
\frac{d}{dx}(\sin x) &= \cos x & \frac{d}{dx}(\csc x) &= -\csc x \cot x \\
\frac{d}{dx}(\tan x) &= \sec^2 x & \frac{d}{dx}(\cot x) &= -\csc^2 x
\end{align*}
\]

Note that the Sum and Difference Rules can be applied to sums and differences, respectively, of not just two functions but any finite (integer) number of functions. For example, for three differentiable functions \( f_1, f_2, \) and \( f_3, \)

\[
\frac{d}{dx}(f_1 + f_2 + f_3) = \frac{df_1}{dx} + \frac{df_2}{dx} + \frac{df_3}{dx} \quad \text{by the Sum Rule}
\]

Continuing like this for four functions, then five functions, and so forth, the Sum and Difference Rules combined with the Constant Multiple Rule yield the following formula:
For \( n \geq 1 \) differentiable functions \( f_1, \ldots, f_n \) and constants \( c_1, \ldots, c_n \):

\[
\frac{d}{dx} (c_1 f_1 + \cdots + c_n f_n) = c_1 \frac{df_1}{dx} + \cdots + c_n \frac{df_n}{dx}
\]

Note that the above formula includes differences, by using negative constants. The formula also shows that differentiation is a linear operation, which makes \( \frac{d}{dx} \) a linear operator. The idea is that \( \frac{d}{dx} \) “operates” on differentiable functions by taking their derivatives with respect to the variable \( x \). The sum \( c_1 f_1 + \cdots + c_n f_n \) is called a linear combination of functions, and the derivative of that linear combination can be taken term by term, with the constant multiples taken outside the derivatives.

A special case of the above formula is replacing the functions \( f_1, \ldots, f_n \) by nonnegative powers of \( x \), making the sum a polynomial in \( x \). In previous sections the derivatives of a few polynomials—such as \( x \) and \( x^2 \)—were calculated. For the derivative of a general polynomial, the following rule is needed:

**Power Rule:** \[
\frac{d}{dx} (x^n) = n x^{n-1} \quad \text{for any integer } n
\]

There are several ways to prove this formula; one such way being a proof by induction, which in general means using the following principle:

**Principle of Mathematical Induction**
A statement \( P(n) \) about integers \( n \geq k \) is true for all \( n \geq k \) if:

1. \( P(k) \) is true.
2. If \( P(n) \) is true for some integer \( n \geq k \) then \( P(n+1) \) is true.

The idea behind mathematical induction is simple: if a statement is true about some initial integer \( k \) (Step 1 above) and if the statement being true for some integer implies it is true for the next integer (Step 2 above), then the statement being true for \( k \) implies it is true for \( k + 1 \), which in turn implies it is true for \( k + 2 \), which implies it is true for \( k + 3 \), and so forth, making it true for all integers \( n \geq k \).

Typically the initial integer \( k \) will be 0 or 1. To prove the Power Rule for all integers, first use induction to prove the rule for all nonnegative integers \( n \geq 0 \), using \( k = 0 \) for the initial integer. The proof by induction then proceeds as follows:

Let \( P(n) \) be the statement: \( \frac{d}{dx} (x^n) = n x^{n-1} \)

1. Show that \( P(0) \) is true.
   That means showing that the Power Rule holds for \( n = 0 \), i.e. \( \frac{d}{dx} (x^0) = 0 x^{0-1} = 0 \). But \( x^0 = 1 \) is a constant, so its derivative is 0. ✓
2. Assuming \( P(n) \) is true for some \( n \geq 0 \), show that \( P(n + 1) \) is true.

Assuming that \( \frac{d}{dx} (x^n) = nx^{n-1} \), show that \( \frac{d}{dx} (x^{n+1}) = (n+1)x^{(n+1)-1} = (n+1)x^n \). It was shown in Section 1.2 that \( \frac{d}{dx} (x) = 1 \), so:

\[
\frac{d}{dx} (x^{n+1}) = \frac{d}{dx} (x \cdot x^n) = x \cdot \frac{d}{dx} (x^n) + x^n \cdot \frac{d}{dx} (x) \quad \text{(by the Product Rule)}
\]

\[
= x \cdot n x^{n-1} + x^n \cdot 1 \quad \text{(by the assumption that \( P(n) \) is true)}
\]

\[
= nx^n + x^n = (n+1)x^n \quad \checkmark
\]

Thus, by induction, the Power Rule is true for all nonnegative integers \( n \geq 0 \).

To show that the Power Rule is true for all negative integers \( n < 0 \), write \( n = -m \), where \( m \) is positive (namely, \( m = |n| \)). Then:

\[
\frac{d}{dx} (x^n) = \frac{d}{dx} (x^{-m}) = \frac{d}{dx} \left( \frac{1}{x^m} \right) = \frac{x^m \cdot \frac{d}{dx} (1) - 1 \cdot \frac{d}{dx} (x^m)}{(x^m)^2} \quad \text{(by the Quotient Rule)}
\]

\[
= \frac{x^m \cdot 0 - 1 \cdot m \cdot x^{m-1}}{x^{2m}} \quad \text{(by the Power Rule for positive integers)}
\]

\[
= -m x^{-1-2m} - m x^{-m-1} = nx^{n-1} \quad \checkmark
\]

Thus, the Power Rule is true for all integers, which completes the proof. \( \text{QED} \)

**Example 1.7**

Find the derivative of \( f(x) = x^4 - 4x^3 + 6x^2 - 4x + 1 \).

**Solution:** Differentiate the polynomial term by term and use the Power Rule:

\[
\frac{df}{dx} = \frac{d}{dx} (x^4 - 4x^3 + 6x^2 - 4x + 1)
\]

\[
= \frac{d}{dx} (x^4) - 4 \cdot \frac{d}{dx} (x^3) + 6 \cdot \frac{d}{dx} (x^2) - 4 \cdot \frac{d}{dx} (x) + \frac{d}{dx} (1)
\]

\[
= 4x^{4-1} - 4 \cdot 3x^{3-1} + 6 \cdot 2x^{2-1} - 4 \cdot 1 + 0
\]

\[
= 4x^3 - 12x^2 + 12x - 4
\]

In general, the derivative of a polynomial of degree \( n \geq 0 \) is given by:

For any constants \( a_0, \ldots, a_n \) with \( n \geq 0 \):

\[
\frac{d}{dx} (a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0) = na_n x^{n-1} + (n-1)a_{n-1} x^{n-2} + \cdots + 2a_2 x + a_1
\]
A way to remember the Power Rule is: bring the exponent down in front of the variable then reduce the variable’s original exponent by 1. This works even for negative exponents.

**Example 1.8**

Find the derivative of \( f(t) = 3t^{100} - \frac{2}{t^{100}} \).

**Solution:** Differentiate term by term:

\[
\frac{df}{dt} = \frac{d}{dt} \left(3t^{100} - 2t^{-100}\right) = 3 \cdot 100t^{99} - 2 \cdot (-100t^{-101}) = 300t^{99} + \frac{200}{t^{101}}
\]

### Exercises

**A**

For Exercises 1-14, use the rules from this section to find the derivative of the given function.

1. \( f(x) = x^2 - x - 1 \)
2. \( f(x) = x^8 + 2x^4 + 1 \)
3. \( f(x) = \frac{2x^6}{3} - \frac{3}{x^6} \)
4. \( f(x) = \frac{\sin x + \cos x}{4} \)
5. \( f(x) = x \sin x \)
6. \( f(x) = x^2 \tan x \)
7. \( f(x) = \frac{\sin x}{x} \)
8. \( f(x) = \frac{\sin x}{x^2} \)
9. \( f(t) = \frac{2t}{1+t^2} \)
10. \( g(t) = \frac{1-t^2}{1+t^2} \)
11. \( f(x) = \frac{ax+b}{cx+d} \) (\(a, b, c, d\) are constants)
12. \( F(r) = -\frac{Gm_1m_2}{r^2} \) (\(G, m_1, m_2\) are constants)
13. \( A(r) = \pi r^2 \)
14. \( V(r) = \frac{4}{3} \pi r^3 \)
15. Show that \( \frac{d}{dx} (\cot x) = -\csc^2 x \).
16. Show that \( \frac{d}{dx} (\csc x) = -\csc x \cot x \).

**B**

17. Prove the Difference Rule.
18. Prove the Constant Multiple Rule.

19. Use the Product Rule to show that for three differentiable functions \( f, g, \) and \( h \), the derivative of their product is \( (fgh)' = f'gh + fg'h + fgh' \).

**C**

20. Provide an alternative proof of the Product Rule for two differentiable functions \( f \) and \( g \) of \( x \):

   (a) Show that \( (df)(dg) = 0 \).

   (b) By definition, the derivative of the product \( f \cdot g \) is

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

Use that formula along with part (a) to show that \( \frac{d}{dx} (f \cdot g) = f \cdot \frac{dg}{dx} + g \cdot \frac{df}{dx} \).

(Hint: Recall that \( df = f(x+dx) - f(x) \).)
1.5 The Chain Rule

From what has been discussed so far it might be tempting to think that the derivative of a function like \( \sin 2x \) is simply \( \cos 2x \), since the derivative of \( \sin x \) is \( \cos x \). It turns out that is not correct:

\[
\frac{d}{dx}(\sin 2x) = \frac{d}{dx}(2 \sin x \cos x) \quad \text{(by the double-angle formula for sine)}
\]

\[
= 2 \frac{d}{dx}(\sin x \cos x) \quad \text{(by the Constant Multiple Rule)}
\]

\[
= 2 \left( \sin x \cdot \frac{d}{dx}(\cos x) + \cos x \cdot \frac{d}{dx}(\sin x) \right) \quad \text{(by the Product Rule)}
\]

\[
= 2 (\sin x \cdot (-\sin x) + \cos x \cdot \cos x)
\]

\[
= 2 (\cos^2 x - \sin^2 x)
\]

\[
= 2 \cos 2x \quad \text{(by the double-angle formula for cosine)}
\]

So the derivative of \( \sin 2x \) is \( 2 \cos 2x \), not \( \cos 2x \).

In other words, you cannot simply replace \( x \) by \( 2x \) in the derivative formula for \( \sin x \). Instead, regard \( \sin 2x \) as a composition of two functions: the sine function and the \( 2x \) function. That is, let \( f(u) = \sin u \), where the variable \( u \) itself represents a function of another variable \( x \), namely \( u(x) = 2x \). So since \( f \) is a function of \( u \), and \( u \) is a function of \( x \), then \( f \) is a function of \( x \), namely: \( f(x) = \sin 2x \). Since \( f \) is a differentiable function of \( u \), and \( u \) is a differentiable function of \( x \), then \( \frac{df}{du} \) and \( \frac{du}{dx} \) both exist (with \( \frac{df}{du} = \cos u \) and \( \frac{du}{dx} = 2 \)), and multiplying the derivatives shows that \( f \) is a differentiable function of \( x \):

\[
\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx} \quad \text{since the infinitesimals } du \text{ cancel, so}
\]

\[
(\cos u) \cdot 2 = \frac{df}{dx} \quad \Rightarrow \quad \frac{df}{dx} = 2 \cos u = 2 \cos 2x
\]

The above argument holds in general, and is known as the Chain Rule:

**Chain Rule:** If \( f \) is a differentiable function of \( u \), and \( u \) is a differentiable function of \( x \), then \( f \) is a differentiable function of \( x \), and its derivative with respect to \( x \) is:

\[
\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx}
\]

Notice how simple the proof is—the infinitesimals \( du \) cancel.\(^{24}\)

\(^{24}\)Some textbooks give dire warnings to *not* think that \( du \) is an actual quantity that can be canceled. However, you can safely ignore those warnings, because \( du \) is just an infinitesimal and hence can be canceled!
The Chain Rule should make sense intuitively. For example, if \( \frac{df}{du} = 4 \) then that means \( f \) is increasing 4 times as fast as \( u \), and if \( \frac{du}{dx} = 3 \) then \( u \) is increasing 3 times as fast as \( x \), so overall \( f \) should be increasing \( 12 = 4 \cdot 3 \) times as fast as \( x \), exactly as the Chain Rule says.

**Example 1.9**

Find the derivative of \( f(x) = \sin(x^2 + x + 1) \).

*Solution:* The idea is to make a substitution \( u = x^2 + x + 1 \) so that \( f(x) = \sin u \). By the Chain Rule,

\[
\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx} = \frac{d}{du} \sin u \cdot \frac{d}{dx} (x^2 + x + 1) = (\cos u) \cdot (2x + 1) = (2x + 1) \cos(x^2 + x + 1)
\]

after replacing \( u \) by its definition as a function of \( x \) in the last step; the final answer for the derivative should be in terms of \( x \), not \( u \).

In the Chain Rule you can think of the function in question as the composition of an “outer” function \( f \) and an “inner” function \( u \); first take the derivative of the “outer” function then multiply by the derivative of the “inner” function. Think of the “inner” function as a box into which you can put any function of \( x \), and the “outer” function being a function of that empty box.

For instance, for the function \( f(x) = \sin(x^2 + x + 1) \) in the previous example, think of the “outer” function as \( \sin \square \), where \( \square = x^2 + x + 1 \) is the “inner” function, so that

\[
f(x) = \sin(x^2 + x + 1) = \sin \square \]

\[
\frac{df}{dx} = (\cos \square) \cdot \frac{d}{dx} \square = (\cos(x^2 + x + 1)) \cdot \frac{d}{dx} (x^2 + x + 1) = (2x + 1) \cos(x^2 + x + 1)
\]

**Example 1.10**

Find the derivative of \( f(x) = (2x^4 - 3 \cos x)^{10} \).

*Solution:* Here the “outer” function is \( f(\square) = \square^{10} \) and the “inner” function is \( \square = u = 2x^4 - 3 \cos x \):

\[
\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx} = 10 \square^9 \cdot \frac{d}{dx} \square = 10(2x^4 - 3 \cos x)^9 (8x^3 + 3 \sin x)
\]
Recall that the composition $f \circ g$ of two functions $f$ and $g$ is defined as $(f \circ g)(x) = f(g(x))$. Using prime notation the Chain Rule can be written as:

**Chain Rule**: If $g$ is a differentiable function of $x$, and $f$ is a differentiable function on the range of $g$, then $f \circ g$ is a differentiable function of $x$, and its derivative with respect to $x$ is:

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$$

Using the Chain Rule, the Power Rule can be extended to include exponents that are rational numbers:

$$\frac{d}{dx} (x^r) = rx^{r-1} \quad \text{for any rational number } r$$

To prove this, let $r = m/n$, where $m$ and $n$ are integers with $n \neq 0$. Then $y = x^r = x^{m/n} = (x^m)^{1/n}$, so that $y^n = x^m$. Taking the derivative with respect to $x$ of both sides of this equation gives

$$\frac{d}{dx} (y^n) = \frac{d}{dx} (x^m)\quad \text{so evaluating the left side by the Chain Rule gives}$$

$$ny^{n-1} \cdot \frac{dy}{dx} = mx^{m-1}$$

$$n(x^{m/n})^{n-1} \cdot \frac{dy}{dx} = mx^{m-1}$$

$$\frac{dy}{dx} = \frac{mx^{m-1}}{n x^{m-(m/n)}} = \frac{m}{n} x^{m-1-(m-(m/n))} = \frac{m}{n} x^{(m/n)-1} = r x^{r-1}$$

**Example 1.11**

Find the derivative of $f(x) = \sqrt{x}$.

*Solution*: Since $\sqrt{x} = x^{1/2}$ then by the Power Rule:

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

**Example 1.12**

Find the derivative of $f(x) = \frac{2}{3\sqrt{x}}$.

*Solution*:

$$\frac{df}{dx} = \frac{d}{dx} \left(\frac{2}{3x^{1/2}}\right) = \frac{2}{3} \cdot \frac{-1}{2} x^{-3/2} = -\frac{1}{3x^{3/2}}$$

---

25 It will be shown in Chapter 2 how to define any real number as an exponent. The Power Rule extends to that case as well.
A
For Exercises 1-18, find the derivative of the given function.

1. \( f(x) = (1 - 5x)^4 \)
2. \( f(x) = 5(x^3 + x - 1)^4 \)
3. \( f(x) = \sqrt{1 - 2x} \)
4. \( f(x) = (1 - x^2)^{\frac{3}{2}} \)
5. \( f(x) = \frac{\sqrt{x}}{x + 1} \)
6. \( f(x) = \frac{\sqrt{x} + 1}{\sqrt{x} - 1} \)
7. \( f(t) = \left(\frac{1 - t}{1 + t}\right)^4 \)
8. \( f(x) = \left(\frac{x^2 + 1}{x - 1}\right)^6 \)
9. \( f(x) = \sin^2 x \)
10. \( f(x) = \cos \left(\sqrt{x}\right) \)
11. \( f(x) = 3\tan(5x) \)
12. \( f(x) = A \cos(\omega x + \phi) \) \( (A, \omega, \phi \text{ are constants}) \)
13. \( f(x) = \sec(x^2) \)
14. \( f(x) = \sin^2\left(\frac{1}{1-x}\right) + \cos^2\left(\frac{1}{1-x}\right) \)
15. \( L(\beta) = \frac{1}{\sqrt{1 - \beta^2}} \)
16. \( f(x) = \frac{1}{\pi s} \left(1 + \left(\frac{x-l}{s}\right)^2\right)^{-1} \) \( (s, l \text{ are constants}) \)
17. \( f(x) = \cos(\cos x) \)
18. \( f(x) = \sqrt{1 + \sqrt{x}} \)

B
19. In a certain type of electronic circuit\(^{26}\) the overall gain \( A_v \) is given by

\[ A_v = \frac{A_o}{1 - T} \]

where the loop gain \( T \) is a function of the open-loop gain \( A_o \).

(a) Show that

\[ \frac{dA_v}{dA_o} = \frac{1}{1 - T} - \frac{A_o}{(1 - T)^2} \frac{d(1-T)}{dA_o}. \]

(b) In the case where \( T \) is directly proportional to \( A_o \), use part(a) to show that

\[ \frac{dA_v}{dA_o} = \frac{1}{(1 - T)^2}. \]

(Hint: First show that \( A_0 \cdot \frac{d(1-T)}{dA_o} = -T. \))

20. Show that the Chain Rule can be extended to 3 functions: if \( u \) is a differentiable function of \( x \), \( v \) is a differentiable function of \( u \), and \( f \) is a differentiable function of \( v \), then

\[ \frac{df}{dx} = \frac{df}{dv} \cdot \frac{dv}{du} \cdot \frac{du}{dx} \]

so that \( f \) is a differentiable function of \( x \). Notice that the 3 derivatives are linked together as in a chain (hence the name of the rule). The Chain Rule can be extended to any finite number of functions by the above technique.

21. In an internal combustion engine, as a piston moves downward the connecting rod rotates the crank in the clockwise direction, as shown in Figure 1.5.1 below.\(^{27}\)

![Figure 1.5.1](image)

The point \(A\) can only move vertically, causing the point \(B\) to move around a circle of radius \(a\) centered at the point \(O\), which is directly below the point \(A\) and does not move. As the crank rotates it makes an angle \(\theta\) with the line \(OA\). Let \(l = AB\) and \(s = OA\) as in the picture. Assume that all lengths are measured in centimeters, and let the time variable \(t\) be measured in minutes.

(a) Show that \(s = a \cos \theta + \sqrt{l^2 - a^2 \sin^2 \theta}^{1/2}\) for \(0 \leq \theta \leq \pi\).

(b) The mean piston speed is \(\bar{S}_p = 2LN\), where \(L = 2a\) is the piston stroke, and \(N\) is the rotational velocity of the crank, measured in revolutions per minute (rpm). The instantaneous piston velocity is \(S_p = \frac{ds}{dt}\). Let \(R = l/a\). Show that for \(0 \leq \theta \leq \pi\),

\[
\left| \frac{S_p}{\bar{S}_p} \right| = \frac{\pi}{2} \sin \theta \left[ 1 + \frac{\cos \theta}{\left( R^2 - \sin^2 \theta \right)^{1/2}} \right].
\]

1.6 Higher Order Derivatives

The derivative $f'(x)$ of a differentiable function $f(x)$ can be thought of as a function in its own right, and if it is differentiable then its derivative—denoted by $f''(x)$—is the second derivative of $f(x)$ (the first derivative being $f'(x)$). Likewise, the derivative of $f''(x)$ would be the third derivative of $f(x)$, written as $f'''(x)$. Continuing like this yields the fourth derivative, fifth derivative, and so on. In general the $n$-th derivative of $f(x)$ is obtained by differentiating $f(x)$ a total of $n$ times. Derivatives beyond the first are called higher order derivatives.

Example 1.13

For $f(x) = 3x^4$ find $f''(x)$ and $f'''(x)$.

Solution: Since $f'(x) = 12x^3$ then the second derivative $f''(x)$ is the derivative of $12x^3$, namely:

$$f''(x) = 36x^2$$

The third derivative $f'''(x)$ is then the derivative of $36x^2$, namely:

$$f'''(x) = 72x$$

Since the prime notation for higher order derivatives can be cumbersome (e.g. writing 50 prime marks for the fiftieth derivative), other notations have been created:

**Notation for the second derivative of $y = f(x)$:** The following are all equivalent:

$$f''(x), f^{(2)}(x), \frac{d^2y}{dx^2}, \frac{d^2}{dx^2}(f(x)), y'', y^{(2)}, \dot{y}, \ddot{f}(x), \frac{d^2f}{dx^2}, D^2f(x)$$

**Notation for the $n$-th derivative of $y = f(x)$:** The following are all equivalent:

$$f^{(n)}(x), \frac{d^n y}{dx^n}, \frac{d^n}{dx^n}(f(x)), y^{(n)}, \frac{d^n f}{dx^n}, D^n f(x)$$

Notice that the parentheses around $n$ in the notation $f^{(n)}(x)$ indicate that $n$ is not an exponent—it is the number of derivatives to take. The $n$ in the Leibniz notation $\frac{d^n y}{dx^n}$ indicates the same thing, and in general makes working with higher order derivatives easier:

$$\frac{d^2 y}{dx^2} = \frac{d}{dx} \left( \frac{dy}{dx} \right)$$

$$\frac{d^3 y}{dx^3} = \frac{d}{dx} \left( \frac{d^2 y}{dx^2} \right) = \frac{d^2}{dx^2} \left( \frac{dy}{dx} \right)$$

$$\vdots$$

$$\frac{d^n y}{dx^n} = \frac{d}{dx} \left( \frac{d^{n-1} y}{dx^{n-1}} \right) = \frac{d^{n-1}}{dx^{n-1}} \left( \frac{dy}{dx} \right)$$
A natural question to ask is: what do higher order derivatives represent? Recall that the first derivative \( f'(x) \) represents the instantaneous rate of change of a function \( f(x) \) at the value \( x \). So the second derivative \( f''(x) \) represents the instantaneous rate of change of the function \( f'(x) \) at the value \( x \). In other words, the second derivative is a rate of change of a rate of change.

The most famous example of this is for motion in a straight line: let \( s(t) \) be the position of an object at time \( t \) as the object moves along the line. The motion can take two directions, e.g. forward/backward or up/down. Take one direction to represent positive position and the other to represent negative direction, as in the drawing on the right. The \( \text{(instantaneous) velocity} \) \( v(t) \) of the object at time \( t \) is \( s'(t) \), i.e. the first derivative of \( s(t) \). The \textbf{acceleration} \( a(t) \) of the object at time \( t \) is defined as \( v'(t) \), the instantaneous rate of change of the velocity. Thus, \( a(t) = s''(t) \), i.e. acceleration is the second derivative of position. To summarize:

\[
\begin{align*}
  s(t) &= \text{position at time } t \\
  v(t) &= \text{velocity at time } t \\
        &= \frac{ds}{dt} = s'(t) = \dot{s}(t) \\
  a(t) &= \text{acceleration at time } t \\
        &= \frac{dv}{dt} = v'(t) = \dot{v}(t) \\
        &= \frac{d}{dt} \left( \frac{ds}{dt} \right) = \frac{d^2s}{dt^2} = s''(t) = \ddot{s}(t)
\end{align*}
\]

\textbf{Example 1.14}

Ignoring wind and air resistance, the position \( s \) of a ball thrown straight up with an initial velocity of 34 m/s from a starting point 2 m off the ground is given by \( s(t) = -4.9t^2 + 34t + 2 \) at time \( t \) (measured in seconds) with \( s \) measured in meters. Find the velocity and acceleration of the ball at any time \( t \geq 0 \).

\textbf{Solution}: The ball moves in a straight vertical line, first straight up then straight down until it hits the ground. Its velocity \( v(t) \) is

\[
v(t) = \frac{ds}{dt} = -9.8t + 34 \text{ m/s}
\]

while its acceleration \( a(t) \) is

\[
a(t) = \frac{d^2s}{dt^2} = \frac{d}{dt} \left( \frac{ds}{dt} \right) = \frac{d}{dt}(-9.8t + 34) = -9.8 \text{ m/s}^2,
\]

which is the acceleration due to the force of gravity on Earth. Note that time \( t = 0 \) is the time at which the ball was thrown, so that \( v(0) \) is the initial velocity of the ball. Indeed, \( v(0) = -9.8(0) + 34 = 34 \text{ m/s} \), as expected.
Notice in Example 1.14 that the acceleration of the ball is not only constant but also negative. To see why this makes sense, first consider the case where the ball is moving upward. The ball has an initial upward velocity of 34 m/s then slows down to 0 m/s at the instant it reaches its maximum height above the ground. So the velocity is decreasing, i.e. its rate of change—the acceleration—is negative.

The ball reaches its maximum height above the ground when its velocity is zero, that is, when \( v(t) = -9.8t + 34 = 0 \), i.e. at time \( t = 34/9.8 = 3.47 \) seconds after being thrown (see the picture above). The ball then starts moving downward and its velocity is negative (e.g. at time \( t = 4 \) s the velocity is \( v(4) = -9.8(4) + 34 = -5.2 \) m/s). Recall that negative velocity indicates downward motion, while positive velocity means the motion is upward (away from the Earth’s center). So in the case where the ball begins moving downward it goes from 0 m/s to a negative velocity, with the ball moving faster towards the ground, which it hits with a velocity of \(-33.43 \) m/s (why?). So again the velocity is decreasing, which again means that the acceleration is negative.

Common terminology involving motion might cause some confusion with the above discussion. For example, even though the ball’s acceleration is negative as it falls to the ground, it is common to say that the ball is accelerating in that situation, not decelerating (as the ball is doing while moving upward). In general, acceleration is understood to mean that the magnitude (i.e. the absolute value) of the velocity is increasing. That magnitude is called the speed of the object. Deceleration means the speed is decreasing.

The first and second derivatives of an object’s position with respect to time represent the object’s velocity and acceleration, respectively. Do the third, fourth, and other higher order derivatives have any physical meanings? It turns out they do. The third derivative of position is called the jerk of the object. It represents the rate of change of acceleration, and is often used in fields such as vehicle dynamics (e.g. minimizing jerk to provide smoother braking). The fourth, fifth, and sixth derivatives of position are called snap, crackle, and pop, respectively.\(^{28}\)

The zero-th derivative \( f^{(0)}(x) \) of a function \( f(x) \) is defined to be the function \( f(x) \) itself: \( f^{(0)}(x) = f(x) \). There is a way to define fractional derivatives, e.g. the one-half derivative \( f^{(1/2)}(x) \), which will be discussed later.

An immediate consequence of the definition of higher order derivatives is:

\[
\frac{d^{m+n}}{dx^{m+n}} (f(x)) = \frac{d^m}{dx^m} \left( \frac{d^n}{dx^n} (f(x)) \right) \text{ for all integers } m \geq 0 \text{ and } n \geq 0.
\]

Recall that the factorial \( n! \) of an integer \( n > 0 \) is the product of the integers from 1 to \( n \):

\[
n! = 1 \cdot 2 \cdot 3 \cdots \cdot n
\]

\(^{28}\)Yes, those really are their names, obviously inspired by a certain breakfast cereal. Snap has found some uses in flight dynamics, e.g. minimizing snap to optimize flight paths of drones.
Higher Order Derivatives

For example:

\[ 1! = 1 \quad 3! = 1 \cdot 2 \cdot 3 = 6 \]
\[ 2! = 1 \cdot 2 = 2 \quad 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24 \]

By convention 0! is defined to be 1. The following statement can be proved using induction:

\[ \frac{d^n}{dx^n} (x^n) = n! \quad \text{for all integers } n \geq 0 \]

Thus,

\[ \frac{d^{n+1}}{dx^{n+1}} (x^n) = \frac{d}{dx} \left( \frac{d^n}{dx^n} (x^n) \right) = \frac{d}{dx} (n!) = 0 \]

for all integers \( n \geq 0 \), since \( n! \) is a constant. So since any polynomial is just a linear combination of nonnegative powers of a variable (typically \( x \)), then the above result combined with the Sum Rule and the Constant Multiple rule yields this important fact:

The \((n+1)\)-st derivative (“\( n \) plus first derivative”) of a polynomial of degree \( n \) is 0:

For any polynomial \( p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \) of degree \( n \), \( \frac{d^{n+1}}{dx^{n+1}} (p(x)) = 0. \)

This is the basis for the commonly-used statement that “any polynomial can be differentiated to 0” by taking a sufficient number of derivatives. For example, differentiating the polynomial \( p(x) = 100x^{100} + 50x^{99} \) 101 times would yield 0 (as would differentiating more than 101 times).

---

**Exercises**

**A**

For Exercises 1-6 find the second derivative of the given function.

1. \( f(x) = x^3 + x^2 + x + 1 \)
2. \( f(x) = x^2 \sin x \)
3. \( f(x) = \cos 3x \)
4. \( f(x) = \frac{\sin x}{x} \)
5. \( f(x) = \frac{1}{x} \)
6. \( F(r) = \frac{Gm_1m_2}{r^2} \)

7. Find the first five derivatives of \( f(x) = \sin x \). Use those to find \( f^{(100)}(x) \) and \( f^{(2014)}(x) \).
8. Find the first five derivatives of \( f(x) = \cos x \). Use those to find \( f^{(100)}(x) \) and \( f^{(2014)}(x) \).
9. If an object moves along a straight line such that its position \( s(t) \) at time \( t \) is directly proportional to \( t \) for all \( t \) (written as \( s \propto t \)), then show that the object’s acceleration is always 0.

**B**

10. Use induction to show that \( \frac{d^n}{dx^n} (x^n) = n! \) for all integers \( n \geq 1 \).
11. Show that for all integers \( n \geq m \geq 1 \), \( \frac{d^m}{dx^m} (x^n) = \frac{n!}{(n-m)!} x^{n-m} \).
12. Find the general expression for the \( n \)-th derivative of \( f(x) = \frac{1}{ax+b} \) for all constants \( a \) and \( b \) (\( a \neq 0 \)).

13. Show that the function \( y = A \cos(\omega t + \phi) + B \sin(\omega t + \phi) \) satisfies the differential equation
\[
\frac{d^2 y}{dt^2} + \omega^2 y = 0
\]
for all constants \( A, B, \omega, \) and \( \phi \).

14. If \( s(t) \) represents the position at time \( t \) of an object moving along a straight line, then show that:
- \( s' \) and \( s'' \) have the same sign \( \Rightarrow \) the object is accelerating
- \( s' \) and \( s'' \) have opposite signs \( \Rightarrow \) the object is decelerating

15. For all twice-differentiable functions \( f \) and \( g \), show that \((f \cdot g)'' = f'' \cdot g + 2f' \cdot g' + f \cdot g''\).

16. Recall that taking a derivative is a way of operating on a function. That is, think of \( \frac{d}{dx} \) as the differentiation operator on the collection of differentiable functions, taking a function \( f(x) \) to its derivative function \( \frac{df}{dx} \):

\[
\begin{array}{c}
\text{f(x)} \\
\downarrow \frac{d}{dx} \\
\text{df/dx}
\end{array}
\]

Likewise, \( \frac{d^2}{dx^2} \) is an operator on twice-differentiable functions, taking a function \( f(x) \) to its second derivative function \( \frac{d^2f}{dx^2} \):

\[
\begin{array}{c}
\text{f(x)} \\
\downarrow \frac{d^2}{dx^2} \\
\text{d^2f/dx^2}
\end{array}
\]

In general, an eigenfunction of an operator \( A \) is a function \( \phi(x) \) such that \( A(\phi(x)) = \lambda \cdot \phi(x) \), that is,

\[
\begin{array}{c}
\phi(x) \\
\downarrow A \\
\lambda \cdot \phi(x)
\end{array}
\]

for all \( x \) in the domain of \( \phi \), for some constant \( \lambda \) called the eigenvalue of the eigenfunction.

(a) Show for all constants \( k \) that \( \phi(x) = \cos(kx) \) is an eigenfunction of the \( \frac{d^2}{dx^2} \) operator, and find its eigenvalue. That is, show that \( \frac{d^2}{dx^2} (\cos(kx)) = \lambda \cdot \cos(kx) \) for some constant \( \lambda \) (the eigenvalue).

(b) The wave function \( \psi \) for a particle of mass \( m \) moving in a one-dimensional box of length \( L \), given by
\[
\psi(x) = \sqrt{2} \frac{\sin(\pi x/L)}{L} \quad \text{for} \quad 0 \leq x \leq L,
\]
is a solution (assuming zero potential energy) of the time-independent Schrödinger equation
\[
-\frac{h^2}{8\pi^2 m} \frac{d^2\psi}{dx^2} = E \psi(x)
\]
where \( h \) is Planck’s constant and \( E \) is a constant that represents the total energy of the wave function. Find an expression for the constant \( E \) in terms of the other constants. Notice that this makes \( \psi(x) \) an eigenfunction of the \( \frac{d^2}{dx^2} \) operator.
CHAPTER 2

Derivatives of Common Functions

2.1 Inverse Functions

The derivatives calculated in the previous chapter were mostly for polynomials and a few trigonometric functions. This chapter will show how to find the derivatives of other types of functions, beginning in this section with inverse functions. The idea here is that if a function is differentiable and has an inverse then that inverse function is also differentiable.

Recall that a function is a rule that assigns a single object \( y \) from one set (the range) to each object \( x \) from another set (the domain). That rule can be written as \( y = f(x) \), where \( f \) is the function (see Figure 2.1.1). There is a simple vertical rule for determining whether a rule \( y = f(x) \) is a function: \( f \) is a function if and only if every vertical line intersects the graph of \( y = f(x) \) in the \( xy \)-coordinate plane at most once (see Figure 2.1.2).

Recall that a function \( f \) is one-to-one (often written as \( 1-1 \)) if it assigns distinct values of \( y \) to distinct values of \( x \). In other words, if \( x_1 \neq x_2 \) then \( f(x_1) \neq f(x_2) \). Equivalently, \( f \) is one-to-one if \( f(x_1) = f(x_2) \) implies \( x_1 = x_2 \). There is a simple horizontal rule for determining whether a function \( y = f(x) \) is one-to-one: \( f \) is one-to-one if and only if every horizontal line intersects the graph of \( y = f(x) \) in the \( xy \)-coordinate plane at most once (see Figure 2.1.3).
If a function \( f \) is one-to-one on its domain, then \( f \) has an inverse function, denoted by \( f^{-1} \), such that \( y = f(x) \) if and only if \( f^{-1}(y) = x \). The domain of \( f^{-1} \) is the range of \( f \).

The basic idea is that \( f^{-1} \) “undoes” what \( f \) does, and vice versa. In other words,

\[
 f^{-1}(f(x)) = x \quad \text{for all } x \text{ in the domain of } f, \text{ and } \\
 f(f^{-1}(y)) = y \quad \text{for all } y \text{ in the range of } f. 
\]

Intuitively it is clear that a function is one-to-one (and hence invertible) when it is either strictly increasing or strictly decreasing; if the function, say, increases and then decreases (as in Figure 2.1.3(b)) then the horizontal rule would be violated around that “turning point.” For differentiable functions a positive derivative means the function is increasing, while a negative derivative means the function is decreasing (this will be proved in Chapter 3).

However, a function can still be one-to-one even if its derivative is zero only at isolated points (i.e. not identically zero over an entire interval of points) and either positive everywhere else or negative everywhere else. For example, the function \( f(x) = x^3 \) has derivative \( f'(x) = 3x^2 \), which is zero only at the isolated point \( x = 0 \) and positive for all other values of \( x \). Clearly \( f \) is one-to-one over the set of all real numbers (why?) and hence it has an inverse function \( x = f^{-1}(y) = \sqrt[3]{y} \) defined for all real numbers \( y \) (i.e. the range of \( f \)).

Thus, having a derivative that is either always positive or always negative is sufficient for a function to be one-to-one but not necessary. Having a nonzero derivative is necessary, though, for the inverse function to be differentiable.

In algebra you learned that \( \frac{a}{b} = \frac{1}{\frac{b}{a}} \) for all real numbers \( a \neq 0 \) and \( b \neq 0 \) (and hence \( \frac{b}{a} \neq 0 \)). The same holds true for the infinitesimals \( dy \) and \( dx \) (nonzero by definition) since they can be treated like numbers, which immediately yields a formula for the derivative of an inverse function:

**Derivative of an Inverse Function:** If \( y = f(x) \) is differentiable and has an inverse function \( x = f^{-1}(y) \), then \( f^{-1} \) is differentiable and its derivative is

\[
 \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} \quad \text{if } \frac{dy}{dx} \neq 0.
\]

The inverse of a function would still exist at a point where \( \frac{dy}{dx} = 0 \) but it would not be differentiable there, since its derivative would be the undefined quantity \( \frac{1}{0} \).
Inverse Functions

Since \( y \) is a function of \( x \), \( \frac{dy}{dx} \) will be in terms of \( x \) and hence \( \frac{1}{\frac{dy}{dx}} \) will be in terms of \( x \). However, since (by invertibility) \( x \) is a function of \( y \), \( \frac{dx}{dy} \) would normally be in terms of \( y \), not \( x \), so that the two sides of the equation \( \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} \) are not in the same terms! One way to handle this discrepancy is to use the formula \( y = f(x) \) to solve for \( x \) in terms of \( y \) then substitute that expression into \( \frac{dx}{dy} \), so that \( \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} \) is now in terms of \( y \). That might not always be possible, however (e.g. try solving for \( x \) in the formula \( y = x \sin x \)).

**Example 2.1**

Find the inverse \( f^{-1} \) of the function \( f(x) = x^3 \) then find the derivative of \( f^{-1} \).

**Solution:** The function \( y = f(x) = x^3 \) is one-to-one over the set of all real numbers (why?) so it has an inverse function \( x = f^{-1}(y) \) defined for all real numbers, namely \( x = f^{-1}(y) = \sqrt[3]{y} \).

The derivative of \( f^{-1} \) is
\[
\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{3x^2},
\]
which is in terms of \( x \), so putting it in terms of \( y \) yields
\[
= \frac{1}{3(\sqrt[3]{y})^2} = \frac{1}{3y^{2/3}}
\]
which agrees with the derivative obtained by differentiating \( x = \sqrt[3]{y} \) directly. Note that this derivative is defined for all \( y \) except \( y = 0 \), which occurs when \( x = \sqrt[3]{0} = 0 \), i.e. at the point \( (x, y) = (0, 0) \).

Functions are often expressed in terms of \( x \), so it is common to see an inverse function also expressed in terms of \( x \): writing the inverse of \( f(x) = x^3 \) as \( f^{-1}(x) = \sqrt[3]{x} \) (not as \( f^{-1}(y) = \sqrt[3]{y} \)), as confusing as that might be. In that case, the idea is to switch the roles of \( x \) and \( y \) in the original function \( y = f(x) \), making it \( x = f(y) \), and then write \( y = f^{-1}(x) \) and use
\[
\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}
\]
to put the derivative of \( f^{-1} \) in terms of \( x \), following the same procedure mentioned earlier.

**Example 2.2**

Find the inverse \( f^{-1} \) of the function \( f(x) = x^3 \) then find the derivative of \( f^{-1} \).

**Solution:** Rewrite \( y = f(x) = x^3 \) as \( x = f(y) = y^3 \), so that its inverse function \( y = f^{-1}(x) = \sqrt[3]{x} \) has derivative
\[
\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{3y^2},
\]
which is in terms of \( y \), so putting it in terms of \( x \) yields
\[
= \frac{1}{3(\sqrt[3]{x})^2} = \frac{1}{3x^{2/3}}
\]
which agrees with the derivative obtained by differentiating \( y = \sqrt[3]{x} \) directly.
To obtain a formula in prime notation for the derivative of an inverse function, notice that for all \( x \) in the domain of an invertible differentiable function \( f \),

\[
f^{-1}(f(x)) = x \implies \frac{d}{dx} (f^{-1}(f(x))) = \frac{d}{dx} (x) \implies (f^{-1})'(f(x)) \cdot f'(x) = 1
\]

by the Chain Rule, and hence:

\[
(f^{-1})'(f(x)) = \frac{1}{f'(x)} \text{ if } f'(x) \neq 0
\]

Two equivalent ways to write this are:

\[
(f^{-1})'(c) = \frac{1}{f'(a)} \text{ where } c = f(a) \text{ and } f'(a) \neq 0
\]

and

\[
(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))} \text{ if } f'(f^{-1}(x)) \neq 0
\]

## Exercises

### A

For Exercises 1-8, show that the given function \( y = f(x) \) is one-to-one over the given interval, then find the formulas for the inverse function \( f^{-1} \) and its derivative. Use Example 2.2 as a guide, including putting \( f^{-1} \) and its derivative in terms of \( x \).

1. \( f(x) = x \), for all \( x \)
2. \( f(x) = 3x \), for all \( x \)
3. \( f(x) = x^2 \), for all \( x \geq 0 \)
4. \( f(x) = \sqrt{x} \), for all \( x \geq 0 \)
5. \( f(x) = \frac{1}{x} \), for all \( x > 0 \)
6. \( f(x) = \frac{1}{x} \), for all \( x < 0 \)
7. \( f(x) = \frac{1}{x^2} \), for all \( x > 0 \)
8. \( f(x) = x^5 \), for all \( x \)

9. The unit circle \( x^2 + y^2 = 1 \) does not define \( y \) as a single function of \( x \), since \( y = \pm \sqrt{1-x^2} \) defines two separate functions. But the part of the unit circle in the first quadrant, i.e. for \( 0 \leq x \leq 1 \) and \( 0 \leq y \leq 1 \), does define \( y = f(x) = \sqrt{1-x^2} \) as a single function of \( x \) that is one-to-one on the interval \([0, 1] \). Find the formulas for its inverse function \( f^{-1} \) and its derivative.

### B

10. Show that if \( f \) is differentiable and invertible, and if \( f^{-1} \) is twice-differentiable, then

\[
(f^{-1})''(x) = -\frac{f''(f^{-1}(x))}{(f'(f^{-1}(x)))^3}.
\]
2.2 Trigonometric Functions and Their Inverses

The graphs of the six trigonometric functions are shown in Figure 2.2.1:

Recall that $\sin x$, $\cos x$, $\csc x$, and $\sec x$ have a period of $2\pi$ (i.e. repeat the same values every $2\pi$ radians), while $\tan x$ and $\cot x$ have a period of $\pi$.

The derivatives of the six trigonometric functions—given in Section 1.4—are:

\[
\frac{d}{dx}(\sin x) = \cos x \quad \frac{d}{dx}(\csc x) = -\csc x \cot x
\]
\[
\frac{d}{dx}(\cos x) = -\sin x \quad \frac{d}{dx}(\sec x) = \sec x \tan x
\]
\[
\frac{d}{dx}(\tan x) = \sec^2 x \quad \frac{d}{dx}(\cot x) = -\csc^2 x
\]

The six trigonometric functions are not one-to-one over their entire domains, but recall from trigonometry that they are one-to-one when restricted to smaller domains, and hence have inverse functions, called the inverse trigonometric functions.
For example, $y = \sin x$ is one-to-one over the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$, as shown in Figure 2.2.2 below:

Similarly, recall that $\cos x$ is one-to-one over $[0, \pi]$, $\tan x$ is one-to-one over $(-\frac{\pi}{2}, \frac{\pi}{2})$, $\csc x$ is one-to-one over $(-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2})$, $\sec x$ is one-to-one over $(0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi)$, and $\cot x$ is one-to-one over $(0, \pi)$. Hence, the inverse trigonometric functions $\sin^{-1} x$, $\cos^{-1} x$, $\tan^{-1} x$, $\csc^{-1} x$, $\sec^{-1} x$ and $\cot^{-1} x$ are defined, with the following domains and ranges:

<table>
<thead>
<tr>
<th></th>
<th>$\sin^{-1} x$</th>
<th>$\cos^{-1} x$</th>
<th>$\tan^{-1} x$</th>
<th>$\csc^{-1} x$</th>
<th>$\sec^{-1} x$</th>
<th>$\cot^{-1} x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>domain</td>
<td>$[-1, 1]$</td>
<td>$[-1, 1]$</td>
<td>$(-\infty, \infty)$</td>
<td>$</td>
<td>x</td>
<td>\geq 1$</td>
</tr>
<tr>
<td>range</td>
<td>$[-\frac{\pi}{2}, \frac{\pi}{2}]$</td>
<td>$[0, \pi]$</td>
<td>$(-\frac{\pi}{2}, \frac{\pi}{2})$</td>
<td>$(-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2})$</td>
<td>$(0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi)$</td>
<td>$(0, \pi)$</td>
</tr>
</tbody>
</table>

The graphs of all six inverse trigonometric functions are shown in Figures 2.2.3 and 2.2.4 below:

---

1 The arc notation $\text{arcsin } x$, $\text{arccos } x$, $\text{arctan } x$, $\text{arcsec } x$, $\text{arccsc } x$, $\text{arccot } x$ is often used in place of $\sin^{-1} x$, $\cos^{-1} x$, $\tan^{-1} x$, $\csc^{-1} x$, $\sec^{-1} x$, $\cot^{-1} x$, respectively.
The derivatives of the six inverse trigonometric functions are:

\[
\begin{align*}
\frac{d}{dx}(\sin^{-1}x) &= \frac{1}{\sqrt{1-x^2}} \quad \text{(for } |x| < 1) \\
\frac{d}{dx}(\csc^{-1}x) &= -\frac{1}{|x|\sqrt{x^2-1}} \quad \text{(for } |x| > 1) \\
\frac{d}{dx}(\cos^{-1}x) &= -\frac{1}{\sqrt{1-x^2}} \quad \text{(for } |x| < 1) \\
\frac{d}{dx}(\sec^{-1}x) &= \frac{1}{|x|\sqrt{x^2-1}} \quad \text{(for } |x| > 1) \\
\frac{d}{dx}(\tan^{-1}x) &= \frac{1}{1+x^2} \\
\frac{d}{dx}(\cot^{-1}x) &= -\frac{1}{1+x^2}
\end{align*}
\]

For the derivative of \(\cos^{-1}x\), recall that \(y = \cos^{-1}x\) is an \textit{angle} between 0 and \(\pi\) radians, defined for \(-1 \leq x \leq 1\). Since \(\cos y = x\) by the definition of \(y\), then \(\frac{dy}{dx} = -\sin y\) and

\[
\sin^2 y = 1 - \cos^2 y = 1 - x^2 \quad \Rightarrow \quad \sin y = \pm \sqrt{1-x^2} = \sqrt{1-x^2}
\]

since \(0 \leq y \leq \pi\) (which means \(\sin y\) must be nonnegative). Thus:

\[
\frac{d}{dx}(\cos^{-1}x) = \frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{-\sin y} = -\frac{1}{\sqrt{1-x^2}} \quad \checkmark
\]

For the derivative of \(\sec^{-1}x\), since \(y = \sec^{-1}x\) is defined for \(|x| \geq 1\), then \(0 \leq y < \pi/2\) for \(x \geq 1\) and \(\pi/2 < y \leq \pi\) for \(x \leq -1\). Recall also that \(\sec y\) and \(\tan y\) are both positive when \(0 < y < \pi/2\) and are both negative when \(\pi/2 < y < \pi\). So in both cases the product \(\sec y \tan y\) is nonnegative, i.e. \(\sec y \tan y = |\sec y \tan y|\). Thus, since \(\sec y = x\) and

\[
1 + \tan^2 y = \sec^2 y \quad \Rightarrow \quad \tan^2 y = \sec^2 y - 1 \quad \Rightarrow \quad \tan y = \pm \sqrt{\sec^2 y - 1} = \pm \sqrt{x^2 - 1}
\]

then for \(|x| > 1\):

\[
\frac{d}{dx}(\sec^{-1}x) = \frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{|\sec y \tan y|} = \frac{1}{|x\sqrt{x^2-1}|} = \frac{1}{|x|\sqrt{x^2-1}} \quad \checkmark
\]
The proofs of the derivative formulas for the remaining inverse trigonometric functions are similar, and are left as exercises.

**Example 2.3**

Find the derivative of the function \( y = 3 \tan(\pi - 2x) \).

**Solution:** By the Chain Rule with \( u = \pi - 2x \), the derivative of \( y = 3 \tan(\pi - 2x) = 3 \tan u \) is:

\[
\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = (3 \sec^2 u) (-2) = -6 \sec^2(\pi - 2x)
\]

**Example 2.4**

Find the derivative of the function \( y = \sin^{-1}(x/4) \).

**Solution:** By the Chain Rule with \( u = x/4 \), the derivative of \( y = \sin^{-1}(x/4) = \sin^{-1}u \) is:

\[
\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \frac{1}{\sqrt{1-u^2}} \cdot \frac{1}{4} = \frac{1}{4\sqrt{1-(x^2/16)}} = \frac{1}{\sqrt{16-x^2}}
\]

**Exercises**

A

For Exercises 1-16, find the derivative of the given function \( y = f(x) \).

1. \( y = \sec^23x \)
2. \( y = \csc(x^2 + 1) \)
3. \( y = \cot3x \)
4. \( y = \cos(\tan x) \)
5. \( y = \tan^{-1}(x/3) \)
6. \( y = \sec^{-1}(x^2 + 1) \)
7. \( y = \cot^{-1}3x \)
8. \( y = \cos^{-1}(\sin x) \)
9. \( y = \cot^{-1}(1/x) \)
10. \( y = \tan^{-1}\sqrt{x} \)
11. \( y = (\sin^{-1}3x)^2 \)
12. \( y = \tan^{-1}1 + \tan^{-1}x \)
13. \( y = \tan^{-1}\frac{x-1}{x+1} \)
14. \( y = x \sin^{-1}(2x+1) \)
15. \( y = x \cot^{-1}x \)
16. \( y = \tan^{-1}\frac{1}{x} + \cot^{-1}x \)

17. Find the derivative of \( y = \sin^{-1}x + \cos^{-1}x \). Explain why no derivative formulas were needed.

B

For Exercises 18-21 prove the given derivative formula.

18. \( \frac{d}{dx} (\sin^{-1}x) = \frac{1}{\sqrt{1-x^2}} \)
19. \( \frac{d}{dx} (\tan^{-1}x) = \frac{1}{1+x^2} \)
20. \( \frac{d}{dx} (\cot^{-1}x) = -\frac{1}{1+x^2} \)
21. \( \frac{d}{dx} (\csc^{-1}x) = -\frac{1}{|x|\sqrt{x^2-1}} \)

22. The Chebyshev polynomials \( T_n(x) = \cos(n \cos^{-1}x) \) are defined for all \(|x| \leq 1\) and \( n = 0, 1, 2, \ldots \).

(a) Show that the Chebyshev polynomials \( T_n(x) \) satisfy the differential equation

\[
(1-x^2)T''_n(x) - x T'_n(x) + n^2 T_n(x) = 0
\]

(b) Find polynomial expressions for \( T_0(x) \), \( T_1(x) \) and \( T_2(x) \).

(c) Show that \( T_{n+1}(x) + T_{n-1}(x) = 2x T_n(x) \) for all \( n \geq 1 \). (Hint: Write \( \theta = \cos^{-1}x \) so that \( \cos \theta = x \))
2.3 The Exponential and Natural Logarithm Functions

Functions of the form $a^x$, where the exponent $x$ varies, are called exponential functions. Unless otherwise noted, assume that $a > 0$ (0$^x$ is just 0, and $(-1)^{1/2}$ is not a real number). You already know how $a^x$ is defined when the exponent $x$ is a rational number (i.e. $x = m/n$ where $m$ and $n$ are integers, $n \neq 0$). But what if $x$ were irrational, such as $\sqrt{2}$? What would $3^{\sqrt{2}}$ mean?

The idea is that $3^{\sqrt{2}} = 1.414213562\ldots$ can be approximated by rational numbers $14/10 = 1.4, 141/100 = 1.41, 1414/1000 = 1.414, 14142/10000 = 1.4142$, and so on, taking larger and larger numerators and denominators in the rational approximations to get more and more decimal places of $\sqrt{2}$. Then $3^{\sqrt{2}}$ would be the number that 3 raised to those rational approximations approaches:

$$3^{\sqrt{2}} = \lim_{m/n \to \sqrt{2}} 3^{m/n}$$

Each quantity $3^{m/n}$ is defined inside the above limit, and as the rational numbers $m/n$ get closer to the value of $\sqrt{2}$ it can be shown the limit of the values of $3^{m/n}$ will exist.\(^2\)

$$3^{1.4} = 3^{14/10} = 4.6553672174608$$
$$3^{1.41} = 3^{141/100} = 4.70696500171657$$
$$3^{1.414} = 3^{1414/1000} = 4.72769503526854$$
$$3^{1.4142} = 3^{14142/10000} = 4.72873393017119$$
$$3^{1.41421} = 3^{141421/100000} = 4.72878588090861$$
$$3^{1.414213} = 3^{1414213/1000000} = 4.72880146624114$$

Of course you would never do all this by hand—you would simply use a computer or calculator, which use much more efficient algorithms for calculating powers in general.\(^3\)

All the usual rules of exponents that you learned in algebra apply to $a^x$ when defined in the manner described above, with $a > 0$ and $x$ varying over all real numbers. Of all the possible values for the base $a$, the one that appears the most in mathematics, the sciences and engineering is the base $e$, defined as:

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

The approximate value of is $e = 2.71828182845905\ldots$ (often called the Euler number).\(^2\)


\(^3\)For example, to see how square roots and cube roots are calculated see Chapter 2 in FIKE, C.T., *Computer Evaluation of Mathematical Functions*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.
The limit in this definition means that as \( x \) becomes larger—approaching infinity (\( \infty \))—the values of \( (1 + \frac{1}{x})^x \) approach a number, denoted by \( e \). More decimal places for \( e \) can be obtained by making \( x \) sufficiently large.\(^4\) For example, when \( x = 5 \times 10^6 \) the value is 2.718281555200129. For extremely large values of \( x \), that is, when \( x \gg 1 \) (the symbol \( \gg \) means “much larger than”),

\[
e 
\approx \left(1 + \frac{1}{x}\right)^x \Rightarrow e^{1/x} \approx \left(\left(1 + \frac{1}{x}\right)^x\right)^{1/x} = 1 + \frac{1}{x} \Rightarrow (e^{1/x} - 1)x \approx \left(\frac{1}{x}\right)x = 1,
\]

so letting \( h = 1/x \), and noting that \( h = 1/x \to 0 \) if and only if \( x \to \infty \), yields the useful limit:\(^5\)

\[
\lim_{h \to 0} \frac{e^h - 1}{h} = 1 \quad (2.2)
\]

Using the above limit, the derivative of \( y = e^x \) can be found:

\[
\frac{d}{dx} (e^x) = e^x
\]

**Proof:** Using the limit definition of the derivative for \( f(x) = e^x \),

\[
\frac{d}{dx} (e^x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{e^{x+h} - e^x}{h}
\]

\[
= \lim_{h \to 0} \frac{e^x(e^h - 1)}{h} = e^x \cdot \lim_{h \to 0} \frac{e^h - 1}{h} \quad \text{(since } e^x \text{ does not depend on } h) \]

\[
= e^x \cdot 1 = e^x
\]

QED

In general, for a differentiable function \( u = u(x) \) as the exponent the Chain Rule yields:

\[
\frac{d}{dx} (e^u) = e^u \cdot \frac{du}{dx}
\]

### Example 2.5

Find the derivative of \( y = 4e^{-x^2} \).

**Solution:** \( \frac{dy}{dx} = 4e^{-x^2} \cdot \frac{d}{dx} (-x^2) = 4e^{-x^2} \cdot (-2x) = -8xe^{-x^2} \)

---

\(^4\)It will be shown in Chapter 9 that this limit does in fact exist.

\(^5\)This is admittedly a “hand waving” argument. In Chapter 3 a more exact method will be discussed for proving limits like this one.
The function $e^x$ is often referred to simply as the exponential function, even though there are obviously many exponential functions. What makes the base $e$ so special? Take $y = Ae^{kt}$ to represent the amount of some physical quantity at time $t$, for some constants $A$ and $k$. Then
\[
\frac{dy}{dt} = \frac{d}{dt} (Ae^{kt}) = k \cdot Ae^{kt} = ky,
\]
which says that the instantaneous rate of change of the quantity is directly proportional to the amount present at that instant. It turns out that many physical quantities exhibit that behavior, some of which will be discussed shortly. Conversely, in Chapter 5 it will be shown that any solution to the differential equation $\frac{dy}{dt} = ky$ must be of the form $y = Ae^{kt}$ for some constant $A$. This is what gives the exponential function its special significance.

Let $f(x) = e^x$ be the exponential function. Then $f(x) > 0$ for all $x$ and $f'(x) = f(x) = e^x > 0$ for all $x$, and so $f(x)$ is strictly increasing. The graph is shown in Figure 2.3.1.

![Figure 2.3.1](image1.png)  
**Figure 2.3.1** The exponential function

![Figure 2.3.2](image2.png)  
**Figure 2.3.2** Natural logarithm function

Thus, the exponential function is one-to-one over the set of all real numbers and hence has an inverse function, called the natural logarithm function, denoted (as a function of $x$) as $f^{-1}(x) = \ln x$. The graph is shown in Figure 2.3.2. Below is a summary of the relationship between $e^x$ and $\ln x$:

<table>
<thead>
<tr>
<th>Domain of $\ln x$</th>
<th>All $x &gt; 0$</th>
<th>Range of $e^x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of $\ln x$</td>
<td>All $x$</td>
<td>Domain of $e^x$</td>
</tr>
</tbody>
</table>

\[
y = e^x \text{ if and only if } x = \ln y \quad (2.3)
\]
\[
e^{\ln x} = x \text{ for all } x > 0 \quad (2.4)
\]
\[
\ln (e^x) = x \text{ for all } x \quad (2.5)
\]

The reader should be aware that many—if not most—fields outside of mathematics use the notation $\log x$ instead of $\ln x$ for the natural logarithm function.\(^6\)

\(^6\)This text almost used $\log x$ as well, prevented only by the desire for compatibility with other mathematics texts.
From algebra you should be familiar with the following properties of the natural logarithm, along with their equivalent properties in terms of the exponential function:

\[
\begin{align*}
\ln(ab) &= \ln a + \ln b \\
\ln \left(\frac{a}{b}\right) &= \ln a - \ln b \\
\ln a^b &= b \ln a \\
\ln 1 &= 0
\end{align*}
\]

\[
\begin{align*}
e^a \cdot e^b &= e^{a+b} \\
e^a / e^b &= e^{a-b} \\
(e^a)^b &= e^{ab} \\
e^0 &= 1
\end{align*}
\]

To find the derivative of \( y = \ln x \), use \( x = e^y \):

\[
\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{\frac{1}{e^y}} = \frac{1}{x}
\]

Hence:

\[
\frac{d}{dx} (\ln x) = \frac{1}{x}
\]

In general, for a differentiable function \( u = u(x) \), the Chain Rule yields:

\[
\frac{d}{dx} (\ln u) = \frac{1}{u} \cdot \frac{du}{dx} = \frac{u'}{u}
\]

**Example 2.6**

Find the derivative of \( y = \ln (x^2 + 3x - 1) \).

**Solution:**

\[
\begin{align*}
\frac{dy}{dx} &= \frac{1}{x^2 + 3x - 1} \cdot \frac{d}{dx} (x^2 + 3x - 1) \\
&= \frac{2x + 3}{x^2 + 3x - 1}
\end{align*}
\]

Recall that \(|x| = -x \) for \( x < 0 \), in which case \( \ln(-x) \) is defined and

\[
\frac{d}{dx} (\ln|x|) = \frac{d}{dx} (\ln(-x)) = \frac{1}{-x} \cdot (-1) = \frac{1}{x}.
\]

Note that when using the formula \( \ln \left(\frac{a}{b}\right) = \ln a - \ln b \) in numerical computations—especially on hand-held calculators—it is preferable to use the left side of the equation, i.e. \( \ln \left(\frac{a}{b}\right) \), since the right side \( \ln a - \ln b \) is vulnerable to the problem of subtractive cancellation, which can give an incorrect answer of 0 if \( a \) and \( b \) are nearly equal.

The Exponential and Natural Logarithm Functions • Section 2.3 49

Combine that result with the derivative \( \frac{d}{dx} (\ln x) = \frac{1}{x} \) for \( x > 0 \) to get:

\[
\frac{d}{dx} (\ln |x|) = \frac{1}{x}
\]

Logarithmic Differentiation

For some functions it is easier to differentiate the natural logarithm of the function first and then solve for the derivative of the original function. This technique is called **logarithmic differentiation**, demonstrated in the following two examples.

**Example 2.7**

Find the derivative of \( y = x^x \).

*Solution:* For this example assume \( x > 0 \) (since \( x \) is both the base and the exponent). Note that you cannot use the Power Rule for this function since the exponent \( x \) is a variable, not a fixed number. Instead, take the natural logarithm of both sides of the equation \( y = x^x \) and then take the derivative of both sides and solve for \( y' \):

\[
\ln y = \ln (x^x) = x \cdot \ln x
\]

\[
\frac{d}{dx} (\ln y) = \frac{d}{dx} (x \cdot \ln x)
\]

\[
y' = y \cdot \frac{1 \cdot \ln x + 1}{x} = x^x (\ln x + 1)
\]

**Example 2.8**

Find the derivative of \( y = \frac{(2x+1)^7(3x^3-7x+6)^4}{(1+\sin x)^5} \).

*Solution:* Use logarithmic differentiation by taking the natural logarithm of \( y \) and then use properties of logarithms to simplify the differentiation before solving for \( y' \):

\[
\ln y = \ln \left( \frac{(2x+1)^7(3x^3-7x+6)^4}{(1+\sin x)^5} \right) = \ln ((2x+1)^7(3x^3-7x+6)^4) - \ln ((1+\sin x)^5)
\]

\[
= 7 \ln (2x+1) + 4 \ln (3x^3-7x+6) - 5 \ln (1+\sin x)
\]

\[
\frac{d}{dx} (\ln y) = \frac{d}{dx} \left( 7 \ln (2x+1) + 4 \ln (3x^3-7x+6) - 5 \ln (1+\sin x) \right)
\]

\[
y' = \frac{7}{2x+1} + \frac{4 \cdot 9x^2 - 7}{3x^3 - 7x + 6} - \frac{5 \cos x}{1+\sin x}
\]

\[
y' = y \cdot \left( \frac{14}{2x+1} + \frac{36x^2 - 28}{3x^3 - 7x + 6} - \frac{5 \cos x}{1+\sin x} \right)
\]

\[
= \frac{(2x+1)^7(3x^3-7x+6)^4}{(1+\sin x)^5} \cdot \left( \frac{14}{2x+1} + \frac{36x^2 - 28}{3x^3 - 7x + 6} - \frac{5 \cos x}{1+\sin x} \right)
\]
Radioactive decay

A classic example of the differential equation \( \frac{dy}{dt} = ky \) is the case of exponential decay of a radioactive substance, often referred to simply as radioactive decay. In this case the general solution \( y = Ae^{kt} \) represents the amount of the substance at time \( t \geq 0 \), and the decay constant \( k \) is negative: \( \frac{dy}{dt} < 0 \) since the substance is decaying (i.e. the amount of substance is decreasing) while \( y > 0 \), so \( \frac{dy}{dt} = ky \) implies that \( k < 0 \).

The constant \( A \) is the initial amount of the substance, i.e. the amount at time \( t = 0 \): \( y(0) = Ae^{0t} = Ae^0 = A \). For this reason \( A \) is sometimes denoted by \( A_0 \). The constant \( k \) turns out to be related to the half-life of the substance, defined as the time \( t_H \) required for half the current amount of substance to decay (see Figure 2.3.3).

You might be tempted to think that the half-life is not a constant, that it might change depending on the amount of substance present. For example, perhaps it would take longer for 100 g of the substance to decay to 50 g than it would for 10 g to decay to 5 g. However, this is not so. To see why, pick any \( t \geq 0 \) as the current time, so that \( y(t) = A_0e^{kt} \) is the current amount of the substance. By definition, that amount should be halved when the time \( t_H \) has passed, that is, \( y(t + t_H) = \frac{1}{2}y(t) \). Then \( t_H \) does turn out to be independent of the initial amount \( A_0 \) and depends only on \( k \), since

\[
y(t + t_H) = \frac{1}{2}y(t) \quad \Rightarrow \quad A_0e^{k(t + t_H)} = \frac{1}{2}A_0e^{kt} \quad \Rightarrow \quad A_0e^{kt} \cdot e^{kt_H} = \frac{1}{2}A_0e^{kt}
\]

\[
\Rightarrow \quad e^{kt_H} = \frac{1}{2} \quad \Rightarrow \quad kt_H = \ln\left(\frac{1}{2}\right) = -\ln 2
\]

so that:

\[
t_H = -\frac{\ln 2}{k} \quad \text{and} \quad k = -\frac{\ln 2}{t_H}
\]

Example 2.9

Suppose that 5 mg of a radioactive substance decays to 3 g in 6 hours. Find the half-life of the substance.

Solution: Consider \( A_0 = 5 \) mg as the initial amount, so that \( y(t) = 5e^{kt} \) is the amount at time \( t \geq 0 \) hours. Use the given information that \( y(6) = 3 \) mg to find \( k \), the decay constant of the substance:

\[
3 = y(6) = 5e^{6k} \quad \Rightarrow \quad 6k = \ln\left(\frac{3}{5}\right) \quad \Rightarrow \quad k = \frac{1}{6} \ln 0.6
\]
Then the half-life \( t_H \) is:
\[
t_H = -\frac{\ln 2}{k} = -\frac{\ln 2}{\frac{1}{6}\ln 0.6} = 8.14 \text{ hours}
\]

Note in the above example that the given time \( t = 6 \) was used for finding the constant \( k \) and then the half-life \( t_H \). For the converse problem—given the half-life find the time required for a certain amount to decay—you would do the opposite: use the given \( t_H \) to find \( k \) and then solve for the required time \( t \) from the equation \( y(t) = A_0 e^{kt} \).

**Example 2.10**

Another example of the differential equation \( \frac{dy}{dt} = ky \) is exponential growth of cell bacteria, in which case \( k > 0 \) since the number of cells \( y(t) \) at time \( t \) is increasing.

**Example 2.11**

Another example is for the current \( I \) in a simple series electric circuit with a constant direct current (DC) source of voltage \( V \), a capacitor with capacitance \( C \), a resistor with resistance \( R \), and a switch, as in Figure 2.3.4. If the capacitor is initially uncharged when the switch is open, and if the switch is closed at time \( t = 0 \), then the current \( I(t) \) through the circuit at time \( t \geq 0 \) satisfies (by Kirchoff’s Second Law) the differential equation
\[
RC \frac{dI}{dt} + I = 0 \quad \Rightarrow \quad \frac{dI}{dt} = -I \frac{1}{RC}
\]
so that \( I(t) = I_0 e^{-t/RC} \) where \( I_0 \) is the initial current at \( t = 0 \). Ohm’s Law says that \( V = I_0 R \), so
\[
I(t) = \frac{V}{R} e^{-t/RC}
\]
is the current at time \( t \geq 0 \), which decreases exponentially.

**Example 2.12**

In the previous examples the quantities that decayed or grew exponentially did so as functions of time. There are other possible variables besides time, though. For example, the atmospheric pressure \( p \) measured as a function of height \( h \) above the surface of the Earth satisfies—assuming constant temperature—the differential equation
\[
\frac{dp}{dh} = -\frac{w_0}{p_0} p
\]
where \( p_0 \) is the pressure at height \( h = 0 \) (i.e. ground level) and \( w_0 \) is the weight of a cubic foot of air at pressure \( p_0 \) (with air pressure measured in lbs per square foot and height measured in feet). Thus,
\[
p(h) = p_0 e^{-w_0 p_0 h}.
\]
So the atmospheric pressure decreases exponentially as the height above the ground increases.
Exercises

A

For Exercises 1-12, find the derivative of the given function.

1. \( y = e^{2x} \)  
2. \( y = x e^{x^2} \)  
3. \( y = e^{-x} - e^x \)  
4. \( y = e^{\sin x} \)  
5. \( y = \frac{1 + e^x}{1 - e^x} \)  
6. \( y = \frac{1}{1 + e^{-2x}} \)  
7. \( y = e^{e^x} \)  
8. \( y = e^{2\ln x} \)  
9. \( y = \ln(3x) \)  
10. \( y = \ln(x^2 + 2x + 1)^4 \)  
11. \( y = (\ln(\tan x^2))^3 \)  
12. \( y = \ln(e^x + e^{2x}) \)

13. Show that \( \frac{d}{dx}(\ln(kx)) = \frac{1}{x} \) for all constants \( k > 0 \).

14. Show that \( \frac{d}{dx}(\ln(x^n)) = \frac{n}{x} \) for all integers \( n \geq 1 \).

For Exercises 15-18, use logarithmic differentiation to find \( \frac{dy}{dx} \).

15. \( y = x^{x^2} \)  
16. \( y = x^{\ln x} \)  
17. \( y = x^{\sin x} \)  
18. \( y = \frac{(x + 2)^6(3x - 1)^7}{(1 - 5x)^4} \)

B

19. Suppose it takes 8 hours for 30% of a radioactive substance to decay. Find the half-life of the substance.

20. The radioactive isotope radium-223 has a half-life of 11.43 days. How long would it take for 3 kg of radium-223 to decay to 1 kg?

21. If a certain cell population grows exponentially—i.e. is of the form \( A_0 e^{kt} \) with \( k > 0 \)—and if the population doubles in 6 hours, how long would it take for the population to quadruple?

For Exercises 22-25, use induction to prove the given formula for all \( n \geq 0 \).

22. \( \frac{d^n}{dx^n}(e^{kx}) = k^n e^{kx} \) (any constant \( k \neq 0 \))

23. \( \frac{d^n}{dx^n}(x e^x) = (x + n) e^x \)

24. \( \frac{d^n}{dx^n}(x e^{-x}) = (-1)^n(x - n) e^{-x} \)

25. \( \frac{d^{n+1}}{dx^{n+1}}(x^n \ln x) = \frac{n!}{x} \)

26. Show that \( f'(x) = f(x)(1 - f(x)) \) for the sigmoid neuron function \( f(x) = \frac{1}{1 + e^{-x}} \). This derivative relation is used in neural network learning algorithms.

27. If \( y = C e^{-\kappa t} \cos\left(\sqrt{n^2 - \kappa^2} t + \gamma\right) \) then show that

\[
\frac{d^2y}{dt^2} + 2\kappa \frac{dy}{dt} + n^2 y = 0
\]

for all constants \( C, n, \kappa, \gamma \), with \( 0 \leq \kappa \leq n \).

C

28. Suppose that \( e^y + e^x = e^{y+x} \). Show that \( \frac{dy}{dx} = -e^{y-x} \).

29. For an infinitesimal \( dx \) show that \( e^{dx} = 1 + dx. \) (Hint: Use \( \frac{d}{dx}(e^x) = e^x \).)

30. For an infinitesimal \( dx \) show that \( \ln(1 + dx) = dx \).
2.4 General Exponential and Logarithmic Functions

For a general exponential function \( y = a^x \), with \( a > 0 \), use logarithmic differentiation to find its derivative:

\[
\ln y = \ln (a^x) = x \ln a
\]

\[
\frac{d}{dx} (\ln y) = \frac{d}{dx} (x \ln a) = \ln a
\]

\[
\frac{y'}{y} = \ln a \quad \Rightarrow \quad y' = y \cdot \ln a
\]

Thus, the derivative of \( y = a^x \) is:

\[
\frac{d}{dx} (a^x) = (\ln a) a^x
\]

In general, for an exponent of the form \( u = u(x) \):

\[
\frac{d}{dx} (a^u) = (\ln a) a^u \cdot \frac{du}{dx}
\]

**Example 2.13**

Find the derivative of \( y = 2^{\cos x} \).

*Solution:* This is the case where \( a = 2 \), so:

\[
\frac{dy}{dx} = (\ln 2) 2^{\cos x} \cdot \frac{d}{dx} (\cos x) = -(\ln 2)(\sin x) 2^{\cos x}
\]

Note that any exponential function \( y = a^x \) can be expressed in terms of the exponential function \( e^x \). Since

\[
a^x > 0 \quad \Rightarrow \quad e^{\ln(a^x)} = a^x,
\]

and since \( \ln(a^x) = x \ln a \), then:

\[
a^x = e^{x \ln a}
\]

Computers and calculators often use the above formula to calculate \( a^x \).

The function \( y = a^x \) has an inverse for any \( a > 0 \), except for \( a = 1 \) (in that case \( y = 1^x = 1 \) is just a constant function). To see this, notice that since \( a^x > 0 \) for all \( x \), and \( \ln a < 0 \) for \( 0 < a < 1 \), while \( \ln a > 0 \) for \( a > 1 \), then \( \frac{dy}{dx} = (\ln a) a^x \) is always negative if
0 < a < 1 and always positive if a > 1. Thus, \( y = a^x \) is a strictly decreasing function if 0 < a < 1, and it is a strictly increasing function if a > 1. The graphs in each case are shown in Figure 2.4.1.

\[
\begin{align*}
\text{Figure 2.4.1} & \quad y = a^x \\
\text{Figure 2.4.2} & \quad y = \log_a x
\end{align*}
\]

Hence, for any \( a > 0 \) with \( a \neq 1 \) the function \( f(x) = a^x \) is one-to-one, so it has an inverse function, called the base \( a \) logarithm and denoted by \( f^{-1}(x) = \log_a x \). It is often spoken as “log base \( a \) of \( x \)”. The graphs for \( a < 1 \) and \( a > 1 \) are shown in Figure 2.4.2. Note that the natural logarithm is just the base \( e \) logarithm in the special case with \( a = e \), i.e. \( \ln x = \log_e x \). The base \( a \) logarithm has properties similar to those of the natural logarithm (and the corresponding properties of \( a^x \)):

\[
\begin{align*}
\log_a (bc) &= \log_a b + \log_a c & a^b \cdot a^c &= a^{b+c} \\
\log_a \left(\frac{b}{c}\right) &= \log_a b - \log_a c & \frac{a^b}{a^c} &= a^{b-c} \\
\log_a b^c &= c \log_a b & (a^b)^c &= a^{bc} \\
\log_a 1 &= 0 & a^0 &= 1
\end{align*}
\]

Note that \( \log_a x \) can be put in terms of the natural logarithm, since

\[
x = a^{\log_a x} \quad \Rightarrow \quad \ln x = \ln \left( a^{\log_a x} \right) = (\log_a x) \cdot (\ln a)
\]

so dividing the last expression by \( \ln a \) gives:

\[
\log_a x = \frac{\ln x}{\ln a}
\]

The above formula is useful on calculators that do not have a \( \log_a x \) key or function. Taking the derivative of both sides yields:
General Exponential and Logarithmic Functions

\[
\frac{d}{dx} (\log_a x) = \frac{1}{x \ln a}
\]

In general, when taking the logarithm of a function \( u = u(x) \):

\[
\frac{d}{dx} (\log_a u) = \frac{1}{u \ln a} \cdot \frac{du}{dx} = \frac{u'}{u \ln a}
\]

**Example 2.14**

Find the derivative of \( y = \log_2(\cos 4x) \).

*Solution:* This is the case where \( a = 2 \), so:

\[
\frac{dy}{dx} = \frac{1}{(\cos 4x)(\ln 2)} \cdot \frac{d}{dx}(\cos 4x) = -\frac{4 \sin 4x}{(\ln 2)(\cos 4x)}
\]

The number \( a \) is the **base** of both the logarithm function \( \log_a x \) and the exponential function \( a^x \). Base 2 and base 10 are the most commonly used bases other than base \( e \). Base 10 is how numbers are normally expressed, as combinations of powers of 10 (e.g. \( 2014 = 2 \cdot 10^3 + 0 \cdot 10^2 + 1 \cdot 10^1 + 4 \cdot 10^0 \)). Base 2 is especially useful in computer science, since computers represent all numbers in **binary** format, i.e. as a sequence of zeros and ones, indicating how many successive powers of two to take and then sum up.\(^8\) For example, the number 6 is represented in binary format as 110, since \( 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 4 + 2 + 0 = 6 \).

---

**Exercises**

**A**

For Exercises 1-9, find the derivative of the given function.

1. \( y = \frac{3^x + 3^{-x}}{2} \)
2. \( y = 2^{\ln 3x} \)
3. \( y = 2^{2^x} \)
4. \( y = \tan^{-1} x \)
5. \( y = \log_2 (x^2 + 1) \)
6. \( y = \log_{10} e^x \)
7. \( y = \sin (\log_2 \pi x) \)
8. \( y = \log_2 4^{2x} \)
9. \( y = 8^{\log_2 x} \)

**B**

10. Show that for all constants \( k \) the function \( y = Aa^{\frac{hx}{m}} \) satisfies the differential equation \( \frac{dy}{dx} = ky \). Does this contradict the statement made in Section 2.3 that the only solution to that differential equation is of the form \( y = Ae^{kx} \)? Explain your answer.

---

\(^8\)Binary notation leads to the joke “There are 10 kinds of people in the world: those who understand binary and those who do not.”
3.1 Tangent Lines

Everyone knows that the Earth is not flat, but locally, e.g. in your immediate vicinity, isn’t the Earth effectively flat? In other words, “flat” is a fairly good approximation of the Earth’s surface “near” you, and it simplifies matters enough for you to do some useful things.

This idea of approximating curved shapes by straight shapes is a frequent theme in calculus. Recall from Chapter 1 that by the Microstraightness Property a differentiable curve \( y = f(x) \) actually is a straight line over an infinitesimal interval, having slope \( \frac{dy}{dx} \). The extension of that line to all values of \( x \) is called the tangent line:

For a curve \( y = f(x) \) that is differentiable at \( x = a \), the tangent line to the curve at the point \( P = (a, f(a)) \) is the unique line through \( P \) with slope \( m = f'(a) \). \( P \) is called the point of tangency. The equation of the tangent line is thus given by:

\[
y - f(a) = f'(a) \cdot (x - a)
\]

Figure 3.1.1 on the right shows the tangent line to a curve \( y = f(x) \) at a point \( P \). If you were to look at the curve near \( P \) with a microscope, it would look almost identical to its tangent line through \( P \). Why is this line—of all possible lines through \( P \)—such a good approximation of the curve near \( P \)? It is because at the point \( P \) the tangent line and the curve both have the same rate of change, namely, \( f'(a) \). So the curve’s values and the line’s values change by roughly the same amount slightly away from \( P \) (where the line and curve have the same value), making their values nearly equal.
Example 3.1

Find the tangent line to the curve \( y = x^2 \) at \( x = 1 \).

Solution: By formula (3.1), the equation of the tangent line is

\[
y - f(a) = f'(a)(x - a)
\]

with \( a = 1 \) and \( f(x) = x^2 \). So \( f(a) = f(1) = 1^2 = 1 \). Both the curve \( y = x^2 \) and the tangent line pass through the point \((1, f(1)) = (1, 1)\). The derivative of \( f(x) = x^2 \) is \( f'(x) = 2x \), so \( f'(a) = f'(1) = 2(1) = 2 \), which is the slope of the tangent line at \((1, 1)\). Hence, the equation of the tangent line is \( y - 1 = 2(x - 1) \) or (in slope-intercept form) \( y = 2x - 1 \).

The curve and tangent line are shown in Figure 3.1.2. Near the point \((1, 1)\) the curve and tangent line are close together, but the separation grows farther from that point, especially in the negative \( x \) direction.

In trigonometry you probably learned about tangent lines to circles, where a tangent line is defined as the unique line that touches the circle at only one point, as in the figure on the right. In this case the tangent line is always on one side of the circle, namely, the exterior of the circle; it does not cut through the interior of the circle. In fact, that definition is a special case of the calculus definition. In general, though, the tangent line to any other type of curve will not necessarily be on only one side of the curve, as it was in Example 3.1.

Example 3.2

Find the tangent line to the curve \( y = x^3 \) at \( x = 0 \).

Solution: Use formula (3.1) with \( a = 0 \) and \( f(x) = x^3 \). Then \( f(a) = f(0) = 0^3 = 0 \). The derivative of \( f(x) = x^3 \) is \( f'(x) = 3x^2 \), so \( f'(a) = f'(0) = 3(0)^2 = 0 \). Hence, the equation of the tangent line is \( y - 0 = 0(x - 0) \), which is \( y = 0 \). In other words, the tangent line is the \( x \)-axis itself.

As shown in Figure 3.1.3, the tangent line cuts through the curve. In general it is possible for a tangent line to intersect the curve at more than one point, depending on the function.

Example 3.3

Find the tangent line to the curve \( y = \sin x \) at \( x = 0 \).

Solution: Use formula (3.1) with \( a = 0 \) and \( f(x) = \sin x \). Then \( f(a) = f(0) = \sin 0 = 0 \). The derivative of \( f(x) = \sin x \) is \( f'(x) = \cos x \), so \( f'(a) = f'(0) = \cos 0 = 1 \). Hence, the equation of the tangent line is \( y - 0 = 1(x - 0) \), which is \( y = x \), as in Figure 3.1.4. Near \( x = 0 \), the tangent line \( y = x \) is close to the line \( y = \sin x \), which was shown in Section 1.3 (namely, \( \sin dx = dx \), so that \( \sin x \approx x \) for \( x \ll 1 \)).
There are several important things to note about tangent lines:

- **The slope of a curve’s tangent line is the slope of the curve.**
  Since the slope of a tangent line equals the derivative of the curve at the point of tangency, the **slope of a curve** at a particular point can be defined as the slope of its tangent line at that point. So curves can have varying slopes, depending on the point, unlike straight lines, which have a constant slope. An easy way to remember all this is to think “slope = derivative.”

- **The tangent line to a straight line is the straight line itself.**
  This follows easily from the definition of a tangent line, but is also easy to see with the “slope = derivative” idea: a straight line’s slope (i.e. derivative) never changes, so its tangent line—having the same slope—will be parallel and hence must coincide with the straight line (since they have the points of tangency in common). For example, the tangent line to the straight line $y = -3x + 2$ is $y = -3x + 2$ at every point on the straight line.

- **The tangent line can be thought of as a limit of secant lines.**
  A **secant line** to a curve is a line that passes through two points on the curve. Figure 3.1.5 shows a secant line $L_{PQ}$ passing through the points $P = (x_0, f(x_0))$ and $Q = (w, f(w))$ on the curve $y = f(x)$,

  ![Figure 3.1.5](image)

  which means that as $Q$ approaches $P$ the “limit” of the secant line $L_{PQ}$ has the same slope and goes through the same point $P$ as the tangent line $T_P$.

\[
\lim_{Q\to P} \text{(slope of } L_{PQ}) = \lim_{w\to x_0} \frac{f(w) - f(x_0)}{w - x_0} = f'(x_0) = \text{slope of } T_P
\]
• **Smooth curves have tangent lines, nonsmooth curves do not.** For example, think of the absolute value function \( f(x) = |x| \). Its graph has a sharp edge at the point \((0,0)\), making it nonsmooth there, as shown in Figure 3.1.6(a) below. There is no real way to define a tangent line at \((0,0)\), because as mentioned in Section 1.2, the derivative of \( f(x) \) does not exist at \( x = 0 \). The same holds true for curves with cusps, as in Figure 3.1.6(b).

![Figure 3.1.6](image)

Many lines go through the point of nonsmoothness, some of which are indicated by the dashed lines in the above figures, but none of them can be the tangent line. Sharp edges and cusps have to be “smoothed out” to have a tangent line.

As a point moves along a smooth curve, the corresponding tangent lines to the curve make varying angles with the positive \( x \)-axis—the angle is thus a function of \( x \). Let \( \phi = \phi(x) \) be the smallest angle that the tangent line \( L \) to a curve \( y = f(x) \) makes with the positive \( x \)-axis, so that \(-90° < \phi(x) < 90°\) for all \( x \) (see Figure 3.1.7).

![Figure 3.1.7](image)

As Figure 3.1.7 shows, \(-90° < \phi(x) < 0°\) when the tangent line \( L \) has negative slope, \( 0° < \phi(x) < 90° \) when \( L \) has positive slope, and \( \phi(x) = 0° \) when \( L \) is horizontal (i.e. has zero slope). The slope of a line is usually defined as the rise divided by the run in a right triangle, as shown in the figure on the right. The figure shows as well that by definition of the tangent of an angle, \( \tan \phi(x) \) also equals the rise (opposite) over run (adjacent). Thus, since the slope of \( L \) is \( f'(x) \), this means that \( \tan \phi(x) = f'(x) \). In other words:
The tangent line to a curve \( y = f(x) \) makes an angle \( \phi(x) \) with the positive \( x \)-axis, given by
\[
\phi(x) = \tan^{-1} f'(x).
\]

**Example 3.4**

Find the angle \( \phi \) that the tangent line to the curve \( y = e^{2x} \) at \( x = -\frac{1}{2} \) makes with the positive \( x \)-axis, such that \(-90^\circ < \phi < 90^\circ\).

**Solution:** The angle is \( \phi = \phi(-1/2) = \tan^{-1} f'(-1/2) \), where \( f(x) = e^{2x} \). Since \( f'(x) = 2e^{2x} \), then
\[
\phi = \phi(-1/2) = tan^{-1} f'(-1/2) = tan^{-1} 2e^{-1} = tan^{-1} 0.7358 = 36.3^\circ.
\]

The figure on the right shows the tangent line \( L \) to the curve at \( x = -\frac{1}{2} \) and the angle \( \phi \).

You learned about perpendicular lines in elementary geometry. Figure 3.1.8 shows the natural way to define how a line \( N \) can be perpendicular to a curve \( y = f(x) \) at a point \( P \) on the curve: the line is perpendicular to the tangent line of the curve at \( P \). Call this line \( N \) the **normal line** to the curve at \( P \). Since \( N \) and \( L \) are perpendicular, their slopes are negative reciprocals of each other (provided neither slope is 0). The equation of the normal line follows easily:

The equation of the normal line to a curve \( y = f(x) \) at a point \( P = (a, f(a)) \) is
\[
y - f(a) = -\frac{1}{f'(a)} \cdot (x - a) \quad \text{if} \quad f'(a) \neq 0.
\]

If \( f'(a) = 0 \), then the normal line is vertical and is given by \( x = a \).

**Example 3.5**

Find the normal line to the curve \( y = x^2 \) at \( x = 1 \). (Note: This is the curve from Example 3.1.)

**Solution:** The equation of the normal line is
\[
y - f(a) = -\frac{1}{f'(a)} \cdot (x - a)
\]
with \( a = 1 \), \( f(x) = x^2 \), and \( f'(x) = 2x \). So \( f(a) = 1 \) and \( f'(a) = 2 \). Hence, the equation of the normal line is \( y - 1 = \frac{1}{2} (x - 1) \) or (in slope-intercept form) \( y = -\frac{1}{2} x + \frac{3}{2} \).
**Exercises**

### A

For Exercises 1-12, find the equation of the tangent line to the curve $y = f(x)$ at $x = a$.

1. $f(x) = x^2 + 1$; at $x = 2$
2. $f(x) = x^2 - 1$; at $x = 2$
3. $f(x) = -x^2 + 1$; at $x = 3$
4. $f(x) = 1$; at $x = -1$
5. $f(x) = 4x$; at $x = 1$
6. $f(x) = e^x$; at $x = 0$
7. $f(x) = x^2 - 3x + 7$; at $x = 2$
8. $f(x) = \frac{x + 1}{x - 1}$; at $x = 0$
9. $f(x) = (x^3 + 2x - 1)^3$; at $x = -1$
10. $f(x) = \tan x$; at $x = 0$
11. $f(x) = \sin 2x$; at $x = 0$
12. $f(x) = \sqrt{1 - x^2}$; at $x = 1/\sqrt{2}$

13. Find the equations of the tangent lines to the curve $y = x^3 - 2x^2 + 4x + 1$ which are parallel to the line $y = 3x - 5$.

14. Draw an example of a curve having a tangent line that intersects the curve at more than one point. For Exercises 15-17, find the angle $\phi$ that the tangent line to the curve $y = f(x)$ at $x = a$ makes with the positive $x$-axis, such that $-90^\circ < \phi < 90^\circ$.

15. $f(x) = x^2$; at $x = 2$
16. $f(x) = \cos 2x$; at $x = \pi/6$
17. $f(x) = x^2 + 2x - 3$; at $x = -1$

18. Show that if $\phi(x)$ is the angle that the tangent line to a curve $y = f(x)$ makes with the positive $x$-axis such that $0^\circ \leq \phi(x) < 180^\circ$, then

$$\phi(x) = \begin{cases} \cos^{-1}\left(\frac{1}{\sqrt{1 + (f'(x))^2}}\right) & \text{when } f'(x) \geq 0 \\ \cos^{-1}\left(\frac{-1}{\sqrt{1 + (f'(x))^2}}\right) & \text{when } f'(x) < 0. \end{cases}$$

*Hint: Draw a right triangle.*

For Exercises 19-21, find the angle $\phi$ that the tangent line to the curve $y = f(x)$ at $x = a$ makes with the positive $x$-axis, such that $0^\circ \leq \phi < 180^\circ$.

19. $f(x) = x^2$; at $x = -1$
20. $f(x) = e^{-x}$; at $x = 1$
21. $f(x) = \ln 2x$; at $x = 10$

For Exercises 22-24, find the equation of the normal line to the curve $y = f(x)$ at $x = a$.

22. $f(x) = \sqrt{x}$; at $x = 4$
23. $f(x) = x^2 + 1$; at $x = 2$
24. $f(x) = x^2 - 7x + 4$; at $x = 3$

25. Find the equations of the normal lines to the curve $y = x^3 - 2x^2 - 11x + 3$ which have a slope of $-\frac{1}{3}$.

### B

26. Show that the area of the triangle formed by the $x$-axis, the $y$-axis, and the tangent line to the curve $y = 1/x$ at any point $P$ is constant (i.e. the area is the same for all $P$).

27. For a constant $a > 0$, let $P$ be a point on the curve $y = ax^2$, and let $Q$ be the point where the tangent line to the curve at $P$ intersects the $y$-axis. Show that the $x$-axis bisects the line segment $PQ$.

28. Let $P$ be a point on the curve $y = 1/x$ in the first quadrant, and let $Q$ be the point where the tangent line to the curve at $P$ intersects the $x$-axis. Show that the triangle $\Delta PQ$ is isosceles, where $O$ is the origin.
3.2 Limits: Formal Definition

So far only the intuitive notion of a limit has been used, namely:

A real number $L$ is the limit of $f(x)$ as $x$ approaches $a$ if the values of $f(x)$ can be made arbitrarily close to $L$ by picking values of $x$ sufficiently close to $a$.

That notion can be put in terms of a formal definition as follows:

Let $L$ and $a$ be real numbers. Then $L$ is the limit of a function $f(x)$ as $x$ approaches $a$, written as

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

if for any given number $\epsilon > 0$, there exists a number $\delta > 0$, such that

$$|f(x) - L| < \epsilon \quad \text{whenever} \quad 0 < |x - a| < \delta.$$  

A visual way of thinking of this definition is shown in Figure 3.2.1 below:

![Figure 3.2.1](image)

Figure 3.2.1 says that for any interval around $L$ on the $y$-axis, you will be able to find at least one small interval around $x = a$ (but excluding $a$) on the $x$-axis that the function $y = f(x)$ maps completely inside that interval on the $y$-axis. Choosing smaller intervals around $L$ on the $y$-axis could force you to find smaller intervals around $a$ on the $x$-axis.

In Figure 3.2.1, $f(x)$ is made arbitrarily close to $L$ (within any distance $\epsilon > 0$) by picking $x$ sufficiently close to $a$ (within some distance $\delta > 0$). Since $0 < |x - a| < \delta$ means that $x = a$ itself is excluded, the solid dot at $(a, L)$ could even be a hollow dot. That is, $f(a)$ does not have to equal $L$, or even be defined; $f(x)$ just needs to approach $L$ as $x$ approaches $a$. Thus—perhaps counter-intuitively—the existence of the limit does not actually depend on what happens at $x = a$ itself.
Example 3.6
Show that \( \lim_{x \to a} x = a \) for any real number \( a \).

**Solution:** Though the limit is obvious, the following “epsilon-delta” proof shows how to use the formal definition. The idea is to let \( \epsilon > 0 \) be given, then “work backward” from the inequality \(|f(x) - L| < \epsilon\) to get an inequality of the form \(|x-a| < \delta\), where \( \delta > 0 \) usually depends on \( \epsilon \). In this case the limit is \( L = a \) and the function is \( f(x) = x \), so since
\[
|f(x) - a| < \epsilon \iff |x - a| < \epsilon,
\]
then choosing \( \delta = \epsilon \) means that
\[
0 < |x-a| < \delta \implies |x-a| < \epsilon \implies |f(x) - a| < \epsilon,
\]
which by definition means that \( \lim_{x \to a} x = a \).

Calculating limits in this way might seem silly since—as in Example 3.6—it requires extra effort for a result that is obvious. The formal definition is used most often in proofs of general results and theorems. For example, the rules for limits—listed in Section 1.2—can be proved by using the formal definition.

Example 3.7
Suppose that \( \lim_{x \to a} f(x) \) and \( \lim_{x \to a} g(x) \) both exist. Show that
\[
\lim_{x \to a} (f(x) + g(x)) = \left( \lim_{x \to a} f(x) \right) + \left( \lim_{x \to a} g(x) \right)
\]

**Solution:** Let \( \lim_{x \to a} f(x) = L_1 \) and \( \lim_{x \to a} g(x) = L_2 \). The goal is to show that \( \lim_{x \to a} (f(x) + g(x)) = L_1 + L_2 \). So let \( \epsilon > 0 \). Then \( \epsilon/2 > 0 \), and so by definition there exist numbers \( \delta_1 > 0 \) and \( \delta_2 > 0 \) such that
\[
0 < |x-a| < \delta_1 \implies |f(x) - L_1| < \epsilon/2, \quad \text{and}
0 < |x-a| < \delta_2 \implies |g(x) - L_2| < \epsilon/2.
\]
Now let \( \delta = \min(\delta_1, \delta_2) \). Then \( \delta > 0 \) and
\[
0 < |x-a| < \delta \implies 0 < |x-a| < \delta_1 \quad \text{and} \quad 0 < |x-a| < \delta_2 \implies |f(x) - L_1| < \epsilon/2 \quad \text{and} \quad |g(x) - L_2| < \epsilon/2.
\]
Since \(|A + B| \leq |A| + |B|\) for all real numbers \( A \) and \( B \), then
\[
|f(x) + g(x) - (L_1 + L_2)| = |(f(x) - L_1) + (g(x) - L_2)| \leq |f(x) - L_1| + |g(x) - L_2|
\]
and thus
\[
0 < |x-a| < \delta \implies |f(x) + g(x) - (L_1 + L_2)| < \epsilon/2 + \epsilon/2 = \epsilon
\]
which by definition means that \( \lim_{x \to a} (f(x) + g(x)) = L_1 + L_2 \). √
The proofs of the other limit rules are similar. In general, using the formal definition will not be necessary for evaluating limits of specific functions—in many cases a simple analysis of the function is all that is needed, often from its graph.

**Example 3.8**

Evaluate \( \lim_{x \to 1} f(x) \) for the following function:

\[
  f(x) = \begin{cases} 
    x & \text{if } x > 1 \\
    2 & \text{if } x = 1 \\
    1 & \text{if } x < 1 
  \end{cases}
\]

**Solution:** From the graph of \( f(x) \) in Figure 3.2.2, it is clear that as \( x \) approaches 1 from the right (i.e. for \( x > 1 \)) \( f(x) \) approaches 1 along the line \( y = x \), whereas as \( x \) approaches 1 from the left (i.e. for \( x < 1 \)) \( f(x) \) approaches 1 along the horizontal line \( y = 1 \). Thus, \( \lim_{x \to 1} f(x) = 1 \).

Note that the limit did not depend on the value of \( f(x) \) at \( x = 1 \).

As Example 3.8 shows, what matters for a limit is what happens to the value of \( f(x) \) as \( x \) gets near \( a \), not at \( x = a \) itself. Figure 3.2.3(a) below shows how as \( x \) approaches \( a \), \( f(x) \) approaches a number different from \( f(a) \). Figure 3.2.3(b) shows that \( x = a \) does not even need to be in the domain of \( f(x) \), i.e. \( f(a) \) does not have to be defined. So it will not always be the case that \( \lim_{x \to a} f(x) = f(a) \).

In Example 3.8 the direction in which \( x \) approached the number 1 did not affect the limit. But what if \( f(x) \) had approached different values depending on how \( x \) approached 1? In that case the limit would not exist. The following definitions and notation for one-sided limits will make situations like that simpler to state.

---

Call $L$ the right limit of a function $f(x)$ as $x$ approaches $a$, written as
\[
\lim_{x \to a^+} f(x) = L,
\]
if $f(x)$ approaches $L$ as $x$ approaches $a$ for values of $x$ larger than $a$.

Call $L$ the left limit of a function $f(x)$ as $x$ approaches $a$, written as
\[
\lim_{x \to a^-} f(x) = L,
\]
if $f(x)$ approaches $L$ as $x$ approaches $a$ for values of $x$ smaller than $a$.

The following statement follows immediately from the above definitions:

The limit of a function exists if and only if both its right limit and left limit exist and are equal:

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

Example 3.9

Evaluate $\lim_{x \to 0^-} f(x)$, $\lim_{x \to 0^+} f(x)$, and $\lim_{x \to 0} f(x)$ for the following function:

\[f(x) = \begin{cases} 
x^2 & \text{if } x < 0 \\
2 - x & \text{if } x \geq 0
\end{cases}\]

Solution: From the graph of $f(x)$ in Figure 3.2.4, it is clear that as $x$ approaches 0 from the left (i.e. for $x < 0$) $f(x)$ approaches 0 along the parabola $y = x^2$, whereas as $x$ approaches 0 from the right (i.e. for $x > 0$) $f(x)$ approaches 2 along the line $y = 2 - x$. Hence, $\lim_{x \to 0^-} f(x) = 0$ and $\lim_{x \to 0^+} f(x) = 2$. Thus, $\lim_{x \to 0} f(x)$ does not exist since the left and right limits do not agree at $x = 0$.

Example 3.10

Evaluate $\lim_{x \to 0^+} \sin \left( \frac{1}{x} \right)$.

Solution: For $x > 0$ the function $f(x) = \sin(1/x)$ is defined, and its graph is shown in Figure 3.2.5. As $x$ approaches 0 from the right, $\sin(1/x)$ will be 1 for the numbers $x = 2/\pi$, $2/5\pi$, $2/9\pi$, $2/13\pi$, ... (which approach 0), and $\sin(1/x)$ will be $-1$ for the numbers $x = 2/3\pi$, $2/7\pi$, $2/11\pi$, $2/15\pi$, ... (which also approach 0). So as $x$ approaches 0 from the right, $\sin(1/x)$ will oscillate between 1 and $-1$. Thus, $\lim_{x \to 0^+} \sin \left( \frac{1}{x} \right)$ does not exist.

Figure 3.2.4

Figure 3.2.5
So far only finite limits have been considered, that is, \( L = \lim \limits_{x \to a} f(x) \) where \( L \) is a real (i.e. finite) number. Define an infinite limit, with \( L = \infty \) or \( -\infty \), as follows:

For a real number \( a \), the limit of a function \( f(x) \) equals infinity as \( x \) approaches \( a \), written as
\[
\lim \limits_{x \to a} f(x) = \infty ,
\]
if \( f(x) \) grows without bound as \( x \) approaches \( a \), i.e. \( f(x) \) can be made larger than any positive number by picking \( x \) sufficiently close to \( a \):
For any given number \( M > 0 \), there exists a number \( \delta > 0 \), such that
\[
f(x) > M \quad \text{whenever} \quad 0 < |x - a| < \delta .
\]

For a real number \( a \), the limit of a function \( f(x) \) equals negative infinity as \( x \) approaches \( a \), written as
\[
\lim \limits_{x \to a} f(x) = -\infty ,
\]
if \( f(x) \) grows negatively without bound as \( x \) approaches \( a \), i.e. \( f(x) \) can be made smaller than any negative number by picking \( x \) sufficiently close to \( a \):
For any given number \( M < 0 \), there exists a number \( \delta > 0 \), such that
\[
f(x) < M \quad \text{whenever} \quad 0 < |x - a| < \delta .
\]

The above definitions can be modified accordingly for one-sided limits. If \( \lim \limits_{x \to a} f(x) = \infty \) or \( \lim \limits_{x \to a} f(x) = -\infty \), then the line \( x = a \) is a vertical asymptote of \( f(x) \), and \( f(x) \) approaches the line \( x = a \) asymptotically. The formal definitions are rarely needed.

**Example 3.11**

Evaluate \( \lim \limits_{x \to 0} \frac{1}{x} \).

**Solution:** For \( x \neq 0 \) the function \( f(x) = \frac{1}{x} \) is defined, and its graph is shown in Figure 3.2.6. As \( x \) approaches 0 from the right, \( 1/x \) approaches \( \infty \), that is,
\[
\lim \limits_{x \to 0^+} \frac{1}{x} = \infty .
\]
As \( x \) approaches 0 from the left, \( 1/x \) approaches \( -\infty \), that is,
\[
\lim \limits_{x \to 0^-} \frac{1}{x} = -\infty .
\]
Since the right limit and the left limit are not equal, then \( \lim \limits_{x \to 0} \frac{1}{x} \) does not exist.

Note that the y-axis (i.e. the line \( x = 0 \)) is a vertical asymptote for \( f(x) = \frac{1}{x} \).
Example 3.12

Evaluate \( \lim_{x \to 0} \frac{1}{x^2} \).

Solution: For \( x \neq 0 \) the function \( f(x) = \frac{1}{x^2} \) is defined, and its graph is shown in Figure 3.2.7. As \( x \) approaches 0 from either the right or the left, \( 1/x^2 \) approaches \( \infty \), that is,

\[
\lim_{x \to 0^+} \frac{1}{x^2} = \infty = \lim_{x \to 0^-} \frac{1}{x^2}.
\]

Since the right limit and the left limit both equal \( \infty \), then \( \lim_{x \to 0} \frac{1}{x^2} = \infty \).

Note that the \( y \)-axis (i.e. the line \( x = 0 \)) is a vertical asymptote for \( f(x) = \frac{1}{x^2} \).

In the limit \( \lim_{x \to a} f(x) \) so far only real values of \( a \) have been considered. However, \( a \) could be either \( \infty \) or \( -\infty \):

For a real number \( L \), the limit of a function \( f(x) \) equals \( L \) as \( x \) approaches \( \infty \), written as

\[
\lim_{x \to \infty} f(x) = L,
\]

if \( f(x) \) can be made arbitrarily close to \( L \) for \( x \) sufficiently large and positive:

For any given number \( \epsilon > 0 \), there exists a number \( N > 0 \), such that

\[
|f(x) - L| < \epsilon \text{ whenever } x > N.
\]

For a real number \( L \), the limit of a function \( f(x) \) equals \( L \) as \( x \) approaches \( -\infty \), written as

\[
\lim_{x \to -\infty} f(x) = L,
\]

if \( f(x) \) can be made arbitrarily close to \( L \) for \( x \) sufficiently small and negative:

For any given number \( \epsilon > 0 \), there exists a number \( N < 0 \), such that

\[
|f(x) - L| < \epsilon \text{ whenever } x < N.
\]

The above definitions can be modified accordingly for \( L \) replaced by either \( \infty \) or \( -\infty \). One way to interpret the statement \( \lim_{x \to \infty} f(x) = L \) is: the long-term behavior of \( f(x) \) is to approach a steady-state at \( L \). If \( \lim_{x \to -\infty} f(x) = L \) or \( \lim_{x \to \infty} f(x) = L \), then the line \( y = L \) is a horizontal asymptote of \( f(x) \), and \( f(x) \) approaches the line \( y = L \) asymptotically. Again, for most limits of specific functions, only the intuitive notions are needed.
Example 3.13

From Figures 3.2.6 and 3.2.7, it is clear that
\[ \lim_{x \to \infty} \frac{1}{x} = 0 = \lim_{x \to -\infty} \frac{1}{x} \text{ and } \lim_{x \to \infty} \frac{1}{x^2} = 0 = \lim_{x \to -\infty} \frac{1}{x^2} \]

Note that the x-axis (i.e. the line \( y = 0 \)) is a horizontal asymptote for \( f(x) = \frac{1}{x} \) and \( f(x) = \frac{1}{x^2} \).

Some limits are obvious—you can use them for calculating other limits:

\[
\begin{align*}
\lim_{x \to \infty} x^n &= \begin{cases} 
\infty & \text{for any real } n > 0 \\
0 & \text{for any real } n < 0 
\end{cases} \\
\lim_{x \to -\infty} x^n &= \begin{cases} 
\infty & \text{for } n = 2, 4, 6, 8, \ldots \\
-\infty & \text{for } n = 1, 3, 5, 7, \ldots 
\end{cases} \\
\lim_{x \to \infty} e^x &= \infty & \lim_{x \to -\infty} e^x &= 0 \\
\lim_{x \to \infty} e^{-x} &= 0 & \lim_{x \to -\infty} e^{-x} &= \infty \\
\lim_{x \to \infty} \ln x &= \infty & \lim_{x \to 0^+} \ln x &= -\infty
\end{align*}
\]

A related notion is that of Big O notation (that is the capital letter O, not a zero):

Say that
\[ f(x) = O(g(x)) \text{ as } x \to \infty, \]
spoken as “\( f \) is big O of \( g \)”, if there exist positive numbers \( M \) and \( x_0 \) such that
\[ |f(x)| \leq M|g(x)| \text{ for all } x \geq x_0. \]

For example, obviously \( 2x^3 = O(x^3) \), by picking \( M = 2 \), with \( x_0 \) any positive number. In general, \( f(x) = O(g(x)) \) means that \( f \) exhibits the same long-term behavior as \( g \), up to a constant multiple. You can think of \( g \) as the more basic “type” of function that describes \( f \), as far as long-term behavior.

Example 3.14

Show that \( 5x^4 - 2 = O(x^4) \).

Solution: First, recall from algebra that \( |a + b| \leq |a| + |b| \) for all real numbers \( a \) and \( b \). Thus,
\[ |5x^4 - 2| \leq |5x^4| + |\ -2| = 5|x^4| + 2 \]
for all \( x \). So since \( |x^4| = x^4 \geq 1 \) for all \( x \geq 1 \), then
\[ |5x^4 - 2| \leq 5|x^4| + 2 \leq 5|x^4| + 2|x^4| = |5x^4 - 2| \leq 7|x^4| \text{ for all } x \geq 1, \]
which shows that \( 5x^4 - 2 = O(x^4) \), with \( M = 7 \) and \( x_0 = 1 \).
Some limits need algebraic manipulation before they can be evaluated.

**Example 3.15**

Evaluate \( \lim_{x \to \infty} \left( \sqrt{x+1} - \sqrt{x} \right) \).

**Solution:** Note that both \( \sqrt{x+1} \) and \( \sqrt{x} \) approach \( \infty \) as \( x \) goes to \( \infty \), resulting in a limit of the form \( \infty - \infty \). This is an example of an indeterminate form, which can equal anything (as will be discussed shortly); it does not have to equal 0 (i.e. the \( \infty \)'s do not necessarily “cancel out”). The trick here is to use the conjugate of \( \sqrt{x+1} - \sqrt{x} \), so that

\[
\lim_{x \to \infty} \left( \sqrt{x+1} - \sqrt{x} \right) = \lim_{x \to \infty} \left( \sqrt{x+1} - \sqrt{x} \right) \frac{\sqrt{x+1} + \sqrt{x}}{\sqrt{x+1} + \sqrt{x}} = \lim_{x \to \infty} \frac{(x+1) - x}{\sqrt{x+1} + \sqrt{x}} = \lim_{x \to \infty} \frac{1}{\sqrt{x+1} + \sqrt{x}} = 0
\]

since the numerator is 1 and both terms in the sum in the denominator approach \( \infty \) (i.e. \( \frac{1}{\infty} = 0 \)).

Some other indeterminate forms are \( \infty/\infty \), \( 0/0 \) and \( \infty \cdot 0 \). How would you handle such limits? One way is to use L'Hôpital's Rule\(^2\); a simplified form is stated below:

**L'Hôpital's Rule:** If \( f \) and \( g \) are differentiable functions and

\[
\lim_{x \to a} \frac{f(x)}{g(x)} = \pm \infty \quad \text{or} \quad \frac{0}{0}
\]

then

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

The number \( a \) can be real, \( \infty \), or \( -\infty \).

**Example 3.16**

Evaluate \( \lim_{x \to \infty} \frac{2x - 1}{e^x} \).

**Solution:** This limit is of the form \( \infty/\infty \):

\[
\lim_{x \to \infty} \frac{2x - 1}{e^x} = \frac{\infty}{\infty} = \lim_{x \to \infty} \frac{2}{e^x} \quad \text{by L'Hôpital's Rule}
\]

= 0

since the numerator is 2 and \( e^x \to \infty \) as \( x \to \infty \).

Note that one way of interpreting the limit being 0 is that \( e^x \) grows much faster than \( 2x - 1 \). In fact, using L'Hôpital's Rule it can be shown that \( e^x \) grows much faster than any polynomial, i.e. exponential growth outstrips polynomial growth.

Example 3.17

Evaluate \( \lim_{x \to \infty} \frac{x}{\ln x} \).

Solution: This limit is of the form \( \infty/\infty \):

\[
\lim_{x \to \infty} \frac{x}{\ln x} = \lim_{x \to \infty} \frac{1}{\frac{1}{x}} \quad \text{by L'Hôpital's Rule}
\]

\[
= \lim_{x \to \infty} x = \infty
\]

Note that one way of interpreting the limit being \( \infty \) is that \( x \) grows much faster than \( \ln x \). In fact, using L'Hôpital's Rule it can be shown that any polynomial grows much faster than \( \ln x \), i.e. polynomial growth outstrips logarithmic growth.

Example 3.18

Evaluate \( \lim_{x \to \infty} xe^{-2x} \).

Solution: This limit is of the form \( \infty \cdot 0 \), which can be converted to \( \infty/\infty \):

\[
\lim_{x \to \infty} xe^{-2x} = \lim_{x \to \infty} \frac{x}{e^{2x}} = \lim_{x \to \infty} \frac{1}{2e^{2x}} \quad \text{by L'Hôpital's Rule}
\]

\[
= 0
\]

Note that the limit is another consequence of exponential growth outstripping polynomial growth.

Example 3.19

Evaluate \( \lim_{x \to \infty} \frac{2x^2 - 7x - 5}{3x^2 + 2x - 1} \).

Solution: This limit is of the form \( \infty/\infty \):

\[
\lim_{x \to \infty} \frac{2x^2 - 7x - 5}{3x^2 + 2x - 1} = \lim_{x \to \infty} \frac{4x - 7}{6x + 2} \quad \text{by L'Hôpital's Rule}
\]

\[
= \lim_{x \to \infty} \frac{4}{6} = \frac{2}{3}
\]

Note that the limit ended up being the ratio of the leading coefficients of the polynomials in the numerator and denominator of the original limit. Note also that the lower-order terms (degree less than 2) ended up not mattering. In general you can always discard the lower-order terms when taking the limit of a ratio of polynomials (i.e. a rational function).
Example 3.20
Evaluate \( \lim_{x \to 0} \frac{1 - \cos x}{x} \).

Solution: This limit is of the form 0/0:
\[
\lim_{x \to 0} \frac{1 - \cos x}{x} = 0
\]
\[
= \lim_{x \to 0} \frac{\sin x}{1} \quad \text{by L'Hôpital's Rule}
\]
\[
= \frac{\sin 0}{1} = 0
\]

There is an intuitive justification for L'Hôpital's Rule: since the limit \( \lim_{x \to a} \frac{f(x)}{g(x)} \) uses a ratio to compare how \( f \) changes relative to \( g \) as \( x \) approaches \( a \), then it is really the rates of change of \( f \) and \( g \)—namely \( f' \) and \( g' \), respectively—that are being compared in that ratio, which is what L'Hôpital's Rule says.

The following result provides another way to calculate certain limits:

**Squeeze Theorem**: Suppose that for some functions \( f, g \) and \( h \) there is a number \( x_0 \geq 0 \) such that
\[
g(x) \leq f(x) \leq h(x) \quad \text{for all } x > x_0
\]
and that \( \lim_{x \to \infty} g(x) = \lim_{x \to \infty} h(x) = L \). Then \( \lim_{x \to \infty} f(x) = L \).

Similarly, if \( g(x) \leq f(x) \leq h(x) \) for all \( x \neq a \) in some interval \( I \) containing \( a \), and if \( \lim_{x \to a} g(x) = \lim_{x \to a} h(x) = L \), then \( \lim_{x \to a} f(x) = L \).

Intuitively, the Squeeze Theorem says that if one function is “squeezed” between two functions approaching the same limit, then the function in the middle must also approach that limit. The theorem also applies to one-sided limits \( (x \to a^+ \text{ or } x \to a^-) \).

Example 3.21
Evaluate \( \lim_{x \to \infty} \frac{\sin x}{x} \).

Solution: Since \( -1 \leq \sin x \leq 1 \) for all \( x \), then dividing all parts of those inequalities by \( x > 0 \) yields
\[
-\frac{1}{x} \leq \frac{\sin x}{x} \leq \frac{1}{x} \quad \text{for all } x > 0 \quad \Rightarrow \quad \lim_{x \to \infty} \frac{\sin x}{x} = 0
\]
by the Squeeze Theorem, since \( \lim_{x \to \infty} \frac{-1}{x} = 0 = \lim_{x \to \infty} \frac{1}{x} \).
A

For Exercises 1-18 evaluate the given limit.

1. \( \lim_{x \to 2} \frac{x^2 + 3x - 10}{x^2 - x - 2} \)
2. \( \lim_{x \to \infty} \frac{x^2 + 3x - 10}{2x^2 - x - 2} \)
3. \( \lim_{x \to \infty} \frac{x^2 + 3x - 10}{2x^2 - x - 2} \)
4. \( \lim_{x \to \infty} \frac{x^3 + 3x - 10}{2x^2 - x - 2} \)
5. \( \lim_{x \to \pi/2} \frac{\cos x}{x - \pi/2} \)
6. \( \lim_{x \to \infty} \frac{x^2}{e^x} \)
7. \( \lim_{x \to -\infty} x^2 e^x \)
8. \( \lim_{x \to 0^+} \frac{\ln x}{e^{1/x}} \)
9. \( \lim_{x \to 0} \frac{x - \sin x}{x^3} \)
10. \( \lim_{x \to 0} \frac{\sin 3x}{x} \)
11. \( \lim_{x \to \infty} \frac{\cos x}{x} \)
12. \( \lim_{x \to 0} x \sin \left( \frac{1}{x} \right) \)
13. \( \lim_{x \to 0} \frac{\ln(1 - x) - \sin^2 x}{1 - \cos^2 x} \)
14. \( \lim_{x \to 1} \left( 1 - \frac{1}{\ln x} \right) \)
15. \( \lim_{x \to \pi/2} \sec x - \tan x \)
16. \( \lim_{x \to 0^+} \frac{e^{-1/x}}{x} \)
17. \( \lim_{x \to \infty} \left( \sqrt{x^2 + 4} - x \right) \)
18. \( \lim_{x \to 0} \cot x \)

B

19. The famous “twin paradox,” a result of Einstein’s special theory of relativity, says that if one of a pair of twins leaves the earth in a rocket traveling at a very high speed, then he will be younger than his twin upon returning to earth. This is due to the phenomenon of time dilation, which says that a clock moving with a speed \( v \) relative to a clock at rest in some inertial reference frame counts time slower relative to the clock at rest, by a factor of

\[ \gamma = \frac{1}{\sqrt{1 - \beta^2}} , \]

called the Lorentz factor, where \( \beta = \frac{v}{c} \) is the fraction of the speed of light \( c \) at which the clock is moving \( (c \approx 2.998 \times 10^8 \text{ m/sec}) \). Notice that \( 0 \leq \beta < 1 \) (why?). For example, a clock moving at half the speed of light, so that \( \beta = 0.5 \), would have \( \gamma = 1.1547 \), meaning that the clock runs about 15.47% slower than the clock on earth.

(a) Evaluate \( \lim_{\beta \to 1} \gamma \). What is the physical interpretation of this limit?

(b) Suppose an astronaut and his twin just turned 30 years old when the astronaut leaves earth on a high-speed journey through space. Upon returning to earth the astronaut is 35 and his twin is 70. At roughly what fraction of the speed of light must the astronaut have been traveling?

20. Show that \( \lim_{x \to \infty} \frac{p(x)}{e^x} = 0 \) for all polynomials \( p(x) \) of degree \( n \geq 1 \) with a positive leading coefficient.

21. Show that \( \lim_{x \to \infty} \frac{p(x)}{\ln x} = \infty \) for all polynomials \( p(x) \) of degree \( n \geq 1 \) with a positive leading coefficient.

22. Show that \( 5x^3 + 6x^2 - 4x + 3 = O(x^3) \).

23. Show that \( \frac{2x^2 + 1}{x + 1} = O(x) \). (Hint: Consider \( x \geq 1 \))

24. Call \( h(x) \) an infinitesimal function as \( x \to a \) if \( \lim_{x \to a} h(x) = 0 \). That is, an infinitesimal function approaches zero near some point. Prove the following result, where \( a \) and \( L \) are real numbers:

\[ \lim_{x \to a} f(x) = L \iff f(x) = L + h(x) \text{ for all } x, \text{ where } h(x) \text{ is an infinitesimal function as } x \to a \]
3.3 Continuity

Recall from the previous section that a limit \( \lim_{x \to a} f(x) \) can exist without being equal to \( f(a) \), or with \( f(a) \) not even being defined. Many functions encountered in applications, however, will meet those conditions, and they have a special name:

A function \( f \) is **continuous** at \( x = a \) if

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

Equation (3.2) in the above definition implies that \( f(a) \) is defined, i.e. \( x = a \) is in the domain of \( f \). Figure 3.3.1 below shows some examples of continuity and discontinuity:

In the above figure, \( f \) is not continuous at \( x = x_2 \) because \( \lim_{x \to x_2} f(x) \neq f(x_2) \); \( f \) is not continuous at \( x = x_3 \) because \( \lim_{x \to x_3} f(x) \) does not exist (the right and left limits do not agree)—\( f \) is said to have a **jump discontinuity** at \( x = x_3 \); and \( f \) is not continuous at \( x = x_4 \) because \( f(x_4) \) is not defined. However, \( f \) is continuous at \( x = x_1 \).

A function is continuous if its graph is one unbroken piece over its entire domain. Polynomials, rational functions, trigonometric functions, exponential functions, and logarithmic functions are all continuous on their domains. For example, \( \tan x \) is continuous over its domain, which is broken into disjoint intervals \((-\pi/2, \pi/2), (\pi/2, 3\pi/2), (3\pi/2, 5\pi/2)\), and so forth; the graph is unbroken on each of those intervals. However, \( \tan x \) is not continuous over all of \( \mathbb{R} \), since the function is not defined at all points in \( \mathbb{R} \).

In the language of infinitesimals, a function \( f \) is continuous at \( x = a \) if \( f(a + dx) - f(a) \) is an infinitesimal for any infinitesimal \( dx \). This definition is rarely used.

Physical examples of continuous functions are position, speed, velocity, acceleration, temperature, and pressure. Some discontinuous functions do arise in applications.
Example 3.22

The floor function \( \lfloor x \rfloor \) is defined as
\[ \lfloor x \rfloor = \text{the largest integer less than or equal to } x. \]

In other words, \( \lfloor x \rfloor \) rounds a non-integer down to the previous integer, and integers stay the same. For example, \( \lfloor 0.1 \rfloor = 0 \), \( \lfloor 0.9 \rfloor = 0 \), \( \lfloor 0 \rfloor = 0 \), and \( \lfloor -1.3 \rfloor = -2 \). The graph of \( \lfloor x \rfloor \) is shown in Figure 3.3.2(a).

\[ y = \lfloor x \rfloor \]

(a) Floor function \( \lfloor x \rfloor \)

Figure 3.3.2 Floor and ceiling functions

Similarly, the ceiling function \( \lceil x \rceil \) is defined as
\[ \lceil x \rceil = \text{the smallest integer greater than or equal to } x. \]

In other words, \( \lceil x \rceil \) rounds a non-integer up to the next integer, and integers stay the same. For example, \( \lceil 0.1 \rceil = 1 \), \( \lceil 0.9 \rceil = 1 \), \( \lceil 1 \rceil = 1 \), and \( \lceil -1.3 \rceil = -1 \). The graph of \( \lceil x \rceil \) is shown in Figure 3.3.2(b).

Clearly both \( \lfloor x \rfloor \) and \( \lceil x \rceil \) have jump discontinuities at the integers, but both are continuous at all non-integer values of \( x \). Both functions are also examples of step functions, due to the staircase appearance of their graphs. Step functions are useful in situations when you want to model a quantity that takes only a discrete set of values. For example, in a car with a 4-gear transmission, \( f(x) \) could be the gear the transmission has shifted to while the car travels at speed \( x \). Up to a certain speed the car remains in first gear \( (f = 1) \) and then shifts to second gear \( (f = 2) \) after attaining that speed, then it remains in second gear until reaching another speed, upon which the car then shifts to third gear \( (f = 3) \), and so on. In general, discrete changes in state are often modeled with step functions.

Example 3.23

For an extreme case of discontinuity, consider the function
\[ f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational} \end{cases} \]

This function is discontinuous at every value of \( x \) in \( \mathbb{R} \), since within any positive distance \( \delta \) of a real number \( x \)—no matter how small \( \delta \) is—there will be an infinite number of both rational and irrational
numbers. This is a property of \( \mathbb{R} \). So the value of \( f \) will keep jumping between 0 and 1 no matter how close you get to \( x \). In other words, for any number \( a \) in \( \mathbb{R} \), \( f(a) \) exists but it will never equal \( \lim_{x \to a} f(x) \) because that limit will not exist.

By the various rules for limits, it is straightforward to show that sums, differences, constant multiples, products and quotients of continuous functions are continuous. Likewise a continuous function of a continuous function (i.e. a composition of continuous functions) is also continuous. In addition, a continuous function of a finite limit of a function can be passed inside the limit:

If \( f \) is a continuous function and \( \lim_{x \to a} g(x) \) exists and is finite, then

\[
f\left(\lim_{x \to a} g(x)\right) = \lim_{x \to a} f(g(x)).
\]

The same relation holds for one-sided limits.

The above result is useful in evaluating the indeterminate forms \( 0^0 \), \( \infty^0 \), and \( 1^\infty \). The idea is to take the natural logarithm of the limit by passing the continuous function \( \ln x \) inside the limit and evaluate the resulting limit.

**Example 3.24**

Evaluate \( \lim_{x \to 0^+} x^x \).

**Solution:** This limit is of the form \( 0^0 \), so let \( y = \lim_{x \to 0^+} x^x \) and then take the natural logarithm of \( y \):

\[
\ln y = \ln \left( \lim_{x \to 0^+} x^x \right) = \lim_{x \to 0^+} \ln x^x = \lim_{x \to 0^+} x \ln x \to 0 \cdot (-\infty) = \lim_{x \to 0^+} \frac{\ln x}{1/x} = \lim_{x \to 0^+} \frac{-1/x^2}{-1/x^2} = \lim_{x \to 0^+} (-x) = 0
\]

Thus, \( \lim_{x \to 0^+} x^x = y = e^0 = 1 \).
There is an important relationship between differentiability and continuity:

Every differentiable function is continuous.

Proof: If a function $f$ is differentiable at $x = a$ then $f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$ exists, so

$$\lim_{x \to a} (f(x) - f(a)) = \lim_{x \to a} (f(x) - f(a)) \cdot \frac{x - a}{x - a} = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \cdot \lim_{x \to a} (x - a) = f'(a) \cdot 0 = 0$$

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

Note that the converse is not true. For example, the absolute value function $f(x) = |x|$ is continuous everywhere—its graph is unbroken, as shown in the picture on the right—but recall from Example 1.3 in Section 1.2 that it is not differentiable at $x = 0$. Continuous curves can have sharp edges and cusps, but differentiable curves cannot.

Two other important theorems$^4$ about continuous functions are:

**Extreme Value Theorem**: If $f$ is a continuous function on a closed interval $[a, b]$ then $f$ attains both a maximum value and a minimum value on that interval.

**Intermediate Value Theorem**: If $f$ is a continuous function on a closed interval $[a, b]$ then $f$ attains every value between $f(a)$ and $f(b)$.

Example 3.25

Show that there is a solution to the equation \( \cos x = x \).

Solution: Let \( f(x) = \cos x - x \). Since \( f \) is continuous for all \( x \), in particular it is continuous on \([0, 1]\). So since \( f(0) = 1 > 0 \) and \( f(1) = -0.459698 < 0 \), then by the Intermediate Value Theorem there is a number \( c \) in the open interval \((0, 1)\) such that \( f(c) = 0 \), since 0 is between the values \( f(0) \) and \( f(1) \). Hence, \( \cos c - c = 0 \), which means that \( \cos c = c \). That is, \( x = c \) is a solution of \( \cos x = x \).

Note in the above example that the Intermediate Value Theorem does not tell you how to find the solution, just that the solution exists. To find the solution you can use the bisection method: divide the interval \([0, 1]\) in half and apply the Intermediate Value Theorem to each half-interval to determine which one contains the solution; repeat this procedure on that half-interval, resulting in a smaller interval containing the solution, then repeat the procedure over and over, until you eventually obtain an interval so small that the midpoint of that interval can be taken as the solution.

Listing 3.1 below shows one way of implementing the bisection method for Example 3.25 to find the root of \( f(x) = \cos x - x \), using the Python programming language.

```
import math

def f(x):
    return math.cos(x) - x

def bisect(a, b):
    midpt = (a+b)/2.0
    tol = 1e-15
    if b - a > tol:
        val = f(midpt)
        if val*f(a) < 0:
            bisect(a, midpt)
        elif val*f(b) < 0:
            bisect(midpt, b)
        else:
            print("Root = %.13f" % (midpt))
    else:
        print("Root = %.13f" % (midpt))

bisect(0, 1)
```

Line 8 sets the tolerance to \( 10^{-15} \): the program terminates upon reaching an interval whose length is smaller than that. The output is shown below:

```
Root = 0.7390851332152
```

This is the number obtained by taking the cosine of a number (in radians) repeatedly.
Exercises

A
For Exercises 1-18, indicate whether the given function $f(x)$ is continuous or discontinuous at the given value $x = a$ by comparing $f(a)$ with $\lim_{x \to a} f(x)$.

1. $f(x) = |x|$; at $x = 0$
2. $f(x) = |x - 1|$; at $x = 0$
3. $f(x) = |x|$; at $x = 0$
4. $f(x) = |x|$; at $x = 0.3$
5. $f(x) = |x|$; at $x = 0$
6. $f(x) = |x|$; at $x = 0.5$
7. $f(x) = x - |x|$; at $x = 0$
8. $f(x) = x - |x|$; at $x = 0.1$
9. $f(x) = x - |x|$; at $x = 0$
10. $f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0; \end{cases}$ at $x = 0$
11. $f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0; \end{cases}$ at $x = 0$
12. $f(x) = \begin{cases} x^2 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0; \end{cases}$ at $x = 0$
13. $f(x) = \begin{cases} x + 1 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0; \end{cases}$ at $x = 1$
14. $f(x) = \begin{cases} \sin(x^2) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0; \end{cases}$ at $x = 0$
15. $f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0; \end{cases}$ at $x = 0$
16. $f(x) = \begin{cases} 0 & \text{if } x \text{ is rational}, \\ 1 & \text{if } x \text{ is irrational}; \end{cases}$ at $x = \sqrt{2}$
17. $f(x) = \begin{cases} 0 & \text{if } x \text{ is rational}, \\ x & \text{if } x \text{ is irrational}; \end{cases}$ at $x = 0$
18. $f(x) = \begin{cases} 0 & \text{if } x \text{ is rational}, \\ x & \text{if } x \text{ is irrational}; \end{cases}$ at $x = 1$
19. Evaluate $\lim_{x \to 0^+} x^2$.
20. Evaluate $\lim_{x \to \infty} x^{1/x}$.
21. Evaluate $\lim_{x \to 0} (1 - x)^{1/x}$.
22. If $f(x) = \frac{x^2 + x - 2}{x - 1}$ for $x \neq 1$, how should $f(1)$ be defined so that $f(x)$ is continuous at $x = 1$?
23. If $f(x) = 1/x$ for $x \neq 0$, is there a way to define $f(0)$ so that $f(x)$ is continuous for all $x$?

B
24. Can a function that is not continuous over a closed interval attain a maximum value and a minimum value in that interval? If so, then give an example; if not then explain why.
25. Show that there is a number $x$ such that $x^3 - x = 3$.
26. Prove that $f(x) = x^5 + 3x^4 - 1$ has at least two distinct real roots.
27. Suppose that a function $f$ is continuous on the interval $[0, 3]$, $f$ has no roots in $[0, 3]$, and $f(1) = 1$. Prove that $f(x) > 0$ for all $x$ in $[0, 3]$.
28. Show that an object whose average speed is $v_{avg}$ over the time interval $a \leq t \leq b$ will move with speed $v_{avg}$ at some time $t$ in $[a, b]$.
29. Let $f(x) = 1/(x - 1)$. Then $f(0) = -1 < 0$ and $f(2) = 1 > 0$. Can you conclude by the Intermediate Value Theorem that $f(x)$ must be 0 for some $x$ in $[0, 2]$? Explain.
30. Show that if $f'$ and $f''$ exist and are continuous at $x$ then
$$f''(x) = \lim_{h \to 0} \frac{f(x) - 2f(x-h) + f(x-2h)}{h^2}.$$  
31. Show that if $f'$, $f''$ and $f'''$ exist and are continuous at $x$ then
$$f'''(x) = \lim_{h \to 0} \frac{f(x) - 3f(x-h) + 3f(x-2h) - f(x-3h)}{h^3}.$$
3.4 Implicit Differentiation

A function \( y = f(x) \) is usually given by an explicit formula, such as \( y = x^2 \). It is then straightforward to find \( \frac{dy}{dx} \) using the differentiation rules you have learned so far. But suppose instead that you were given merely an equation involving \( x \) and \( y \), such as

\[
x^3 y^2 e^{\sin(xy)} = x^2 + xy + y^3.
\]

The set of points \((x, y)\) satisfying this equation describes some sort of curve in the xy-plane, but it might not be possible to solve for \( y \) in terms of \( x \)—that is, there might not be an explicit formula for \( y \) as a function of the variable \( x \). So in this case does the derivative \( \frac{dy}{dx} \) even have any meaning, and if so then how would you find it?

It turns out that \( \frac{dy}{dx} \) does make sense in such a case, because equations involving \( x \) and \( y \) such as the one above implicitly define \( y \) in terms of \( x \) in the following sense: you might not get \( y \) itself as an explicit function of \( x \), but you can get the derivative of \( y \) with respect to the variable \( x \). The idea is to take \( \frac{d}{dx} \) of both sides of the equation, then assume that \( y \) really is a function of \( x \) so that you can use the Chain Rule to solve for \( \frac{dy}{dx} \). The example below illustrates this procedure, called implicit differentiation.

**Example 3.26**

Find \( \frac{dy}{dx} \) given the equation \( x^3 + 3x + 2 = y^2 \).

**Solution:** The above equation implicitly defines an elliptic curve, and its graph is shown on the right. This curve is not a function \( y = f(x) \), since it violates the vertical line test, but \( y \) still varies with \( x \). To find \( \frac{dy}{dx} \) take \( \frac{d}{dx} \) of both sides of the equation then solve for \( \frac{dy}{dx} \):

\[
\frac{d}{dx}(x^3 + 3x + 2) = \frac{d}{dx}(y^2)
\]

\[
3x^2 + 3 = 2y \cdot \frac{dy}{dx} \quad \text{by the Chain Rule, so}
\]

\[
\frac{dy}{dx} = \frac{3x^2 + 3}{2y}
\]

At first this might seem unsatisfying—or confusing—since \( \frac{dy}{dx} \) is given in terms of both \( x \) and \( y \). However, the derivative can still be evaluated at specific points \((x, y)\) on the curve, i.e. any \((x, y)\) satisfying the original equation. For example, it is easy to check that \((1, \sqrt{6})\) satisfies the equation \( x^3 + 3x + 2 = y^2 \), so \( \frac{dy}{dx}(1, \sqrt{6}) = \frac{3(1)^2 + 3}{2\sqrt{6}} = \frac{\sqrt{6}}{2} \). Note that \( \frac{dy}{dx} \) is not defined when \( y = 0 \).

Notice that taking the square root of both sides of the original equation does not result in an explicit formula for \( y \), since \( y = \pm \sqrt{x^3 + 3x + 2} \) defines two functions, not just one. The beauty of implicit differentiation is that the derivative \( \frac{dy}{dx} = \frac{3x^2 + 3}{2y} \) calculated above gives you a single expression for the derivative of both those functions.
An algebraic curve is defined as the set of all points \((x, y)\) satisfying a polynomial equation in the variables \(x\) and \(y\), such as \(x^2 - 3xy^4 + 1 = x^5 - y^2\). An elliptic curve is a special case of an algebraic curve, where the polynomial has the specific form \(x^3 + ax + b = y^2\), such as the equation \(x^3 + 3x + 2 = y^2\) from Example 3.26. Elliptic curves have certain properties that have found applications in cryptography.\(^5\)

**Example 3.27**

Find \(\frac{dy}{dx}\) given the equation \(x + y = x^3 + y^3\).

**Solution:** The above equation implicitly defines an algebraic curve and its graph is shown on the right. To find \(\frac{dy}{dx}\) take \(\frac{d}{dx}\) of both sides of the equation then solve for \(\frac{dy}{dx}\):

\[
\frac{d}{dx}(x + y) = \frac{d}{dx}(x^3 + y^3)
\]

\[
1 + \frac{dy}{dx} = 3x^2 + 3y^2 \cdot \frac{dy}{dx} \text{ by the Chain Rule, so}
\]

\[
\frac{dy}{dx} = \frac{3x^2 - 1}{1 - 3y^2}
\]

Notice that the curve consists of an ellipse with a line through it. In fact, that line is \(y = -x\), as can be verified by replacing each instance of \(y\) in the equation \(x + y = x^3 + y^3\) by \(-x\) (resulting in the equation \(0 = 0\)). You might be wondering how \(\frac{dy}{dx}\) is defined at the points where that line intersects the ellipse: is it the slope of the line \(y = -x\) (i.e., \(-1\)), or is it the slope of the tangent line to the ellipse at those points (which would not equal \(-1\))? This is discussed in the exercises.

The graph was created with the free open-source graphing program Gnuplot\(^6\) using the following Gnuplot commands (which give an idea of how to plot implicit functions in general):

```
set size square
set view 0,0
set isosamples 500,500
set contour base
set cntrparam levels discrete 0
unset surface
set grid
unset key
unset ztics
set xlabel 'x'
set ylabel 'y'
f(x,y) = x + y - x**3 - y**3
splot [-3:3][-3:3] f(x,y) lw 3
```

---


\(^6\)Available at http://www.gnuplot.info
Example 3.28

Find the tangent line to the curve \( x^2 + y^2 = 1 \) at the point \((4/5, 3/5)\).

Solution: This curve is the unit circle, shown in the picture on the right.

First use implicit differentiation to find \( \frac{dy}{dx} \):

\[
\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(1) \quad \Rightarrow \quad 2x + 2y \cdot \frac{dy}{dx} = 0 \quad \text{by the Chain Rule}
\]

\[
\Rightarrow \quad \frac{dy}{dx} = -\frac{x}{y}
\]

The slope \( m \) of the tangent line to the curve at \((4/5, 3/5)\) is then

\[
m = \frac{dy}{dx}(4/5, 3/5) = -\frac{4}{3}.
\]

Thus, the equation of the tangent line is

\[
y - \frac{3}{5} = -\frac{4}{3} \left( x - \frac{4}{5} \right).
\]

Exercises

A

For Exercises 1-9, use implicit differentiation to find \( \frac{dy}{dx} \).

1. \( x^3 y - 4xy^2 = y + x^2 \)

2. \( xy = (x+y)^3 \)

3. \( (x+y)^3 = (x-y+1)^2 \)

4. \( x^{2/3} + y^{2/3} = a^{2/3} \)

5. \( (x^2 - y^2)^2 = 2x^2 + y^2 \)

6. \( \frac{x+y}{x-y} = x^2 + y^2 \)

7. \( \cos(xy) = \sin(x^2y^2) \)

8. \( x^3 - x = y^2 \)

9. \( x^3y^2e^{\sin(xy)} = x^2 + xy + y^3 \)

10. In Example 3.28 is it possible to solve the equation \( x^2 + y^2 = 1 \) explicitly for \( y \) in terms of \( x \)? Explain.

11. In Example 3.28 what happens to the tangent line at the point \((1,0)\)? Why does this make sense geometrically?

12. Find the equation of the tangent line to the curve \( x^3 + 3x^2y + y^3 = 8 \) at the point \((2,0)\).

B

13. Find \( \frac{d^2y}{dx^2} \) for the curve \( x^2 + y^2 = 1 \). You may use the results from Example 3.28.

14. Show that at every point \((x_0, y_0)\) on the curve \( y^2 = 4ax \), the equation of the tangent line to the curve is \( y_0y = 2a(x + x_0) \).

15. Show that at every point \((x_0, y_0)\) on the ellipse \( \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \), the equation of the tangent line to the ellipse is \( \frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1 \).

16. Show that at every point \((x_0, y_0)\) on the hyperbola \( \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \), the equation of the tangent line to the hyperbola is \( \frac{xx_0}{a^2} - \frac{yy_0}{b^2} = 1 \).

17. Show that \( \frac{dy}{dx} \) is not defined at the points of intersection of the line and ellipse described by the curve \( x + y = x^3 + y^3 \) from Example 3.27. (Hint: Factor the equation \( x + y = x^3 + y^3 \).

18. Show that the points \( P = (2,4) \) and \( Q = (-31/64, -337/512) \) are on the elliptic curve \( x^3 + 3x + 2 = y^2 \) from Example 3.26, and that the tangent line to the curve at \( P \) also goes through \( Q \).
3.5 Related Rates

If several quantities are related by an equation, then differentiating both sides of that equation with respect to a variable (usually \( t \), representing time) produces a relation between the rates of change of those quantities. The known rates of change are then used in that relation to determine an unknown related rate.

Example 3.29

Suppose that water is being pumped into a rectangular pool at a rate of 60,000 cubic feet per minute. If the pool is 300 ft long, 100 ft wide, and 10 ft deep, how fast is the height of the water inside the pool changing?

Solution:

Let \( V \) be the volume of the water in the pool. Since the volume of a rectangular solid is the product of the length, width, and height of the solid, then

\[
V = (300)(100)h = 30000h \text{ ft}^3
\]

where \( h \) is the height of the water, as in the picture on the right. Both \( V \) and \( h \) are functions of time \( t \) (measured in minutes), and \( \frac{dV}{dt} = 60000 \text{ ft}^3/\text{min} \) was given. The goal is to find \( \frac{dh}{dt} \). Since

\[
\frac{dV}{dt} = \frac{d}{dt}(30000h) = 30000 \frac{dh}{dt}
\]

then

\[
\frac{dh}{dt} = \frac{1}{30000} \frac{dV}{dt} = \frac{1}{30000} \cdot 60000 = 2 \text{ ft/min}.
\]

Example 3.30

Suppose that the angle of inclination from the top of a 100 ft pole to the sun is decreasing at a rate of 0.05 radians per minute. How fast is the length of the pole’s shadow on the ground increasing when the angle of inclination is \( \pi/6 \) radians? You may assume that the pole is perpendicular to the ground.

Solution:

Let \( \theta \) be the angle of inclination and let \( x \) be the length of the shadow, as in the picture on the right. Both \( \theta \) and \( x \) are functions of time \( t \) (measured in minutes), and \( \frac{d\theta}{dt} = -0.05 \text{ rad/min} \) was given (the derivative is negative since \( \theta \) is decreasing). The goal is to find \( \frac{dx}{dt} \) when \( \theta = \pi/6 \), denoted by \( \frac{dx}{dt} \bigg|_{\theta=\pi/6} \) (the vertical bar means “evaluated at” the value of the subscript to the right of the bar). Since

\[
x = 100 \cot \theta \quad \Rightarrow \quad \frac{dx}{dt} = -100 \csc^2 \theta \cdot \frac{d\theta}{dt} = -100 \csc^2 \theta \cdot (-0.05) = 5 \csc^2 \theta
\]

then

\[
\frac{dx}{dt} \bigg|_{\theta=\pi/6} = 5 \csc^2(\pi/6) = 5 (2)^2 = 20 \text{ ft/min}.
\]
Example 3.31

The radius of a right circular cylinder is decreasing at the rate of 3 cm/min, while the height is increasing at the rate of 2 cm/min. Find the rate of change of the volume of the cylinder when the radius is 8 cm and the height is 6 cm.

Solution: Let \( r \), \( h \), and \( V \) be the radius, height, and volume, respectively, of the cylinder. Then \( V = \pi r^2 h \). Since \( \frac{dr}{dt} = -3 \) cm/min and \( \frac{dh}{dt} = 2 \) cm/min, then by the Product Rule:

\[
\frac{dV}{dt} = \frac{d}{dt}(\pi r^2 h) = \left( 2\pi r \cdot \frac{dr}{dt} \right) h + \pi r^2 \cdot \frac{dh}{dt}
\]

\[
\frac{dV}{dt} \bigg|_{r=8, h=6} = 2\pi (8)(-3)(6) + \pi (8^2)(2) = -160\pi \text{ cm}^3/\text{min}
\]

Exercises

A

1. A stone is dropped into still water. If the radius of the circular outer ripple increases at the rate of 4 ft/s, how fast is the area of the circle of disturbed water increasing when the radius is 10 ft?

2. The radius of a sphere decreases at a rate of 3 mm/hr. Determine how fast the volume and surface area of the sphere are changing when the radius is 5 mm.

3. A kite 80 ft above level ground moves horizontally at a rate of 4 ft/s away from the person flying it. How fast is the string being released at the instant when 100 ft of string have been released?

4. A 10-ft ladder is leaning against a wall on level ground. If the bottom of the ladder is dragged away from the wall at the rate of 5 ft/s, how fast will the top of the ladder descend at the instant when it is 8 ft from the ground?

5. A person 6 ft tall is walking at a rate of 6 ft/s away from a light which is 15 ft above the ground. At what rate is the end of the person’s shadow moving along the ground away from the light?

6. An object moves along the curve \( y = x^3 \) in the \( xy \)-plane. At what points on the curve are the \( x \) and \( y \) coordinates of the object changing at the same rate?

7. The radius of a right circular cone is decreasing at the rate of 4 cm/min, while the height is increasing at the rate of 3 cm/min. Find the rate of change of the volume of the cone when the radius is 6 cm and the height is 7 cm.

8. Two boats leave the same dock at the same time, one goes north at 25 mph and the other goes east at 30 mph. How fast is the distance between the boats changing when they are 100 miles apart?

9. Repeat Exercise 8 with the angle between the boats being 110°.

B

10. An angle \( \theta \) changes with time. For what values of \( \theta \) do \( \sin \theta \) and \( \tan \theta \) change at the same rate?

11. Repeat Example 3.30 but with the ground making a 100° angle with the pole to the left of the pole.

C

12. An upright cylindrical tank full of water is tipped over at a constant angular speed. Assume that the height of the tank is at least twice its radius. Show that at the instant the tank has been tipped 45°, water is leaving the tank twice as fast as it did at the instant the tank was first tipped. (Hint: Think of how the water looks inside the tank as it is being tipped.)
3.6 Differentials

An ideal gas satisfies the equation $PV = RT$, where $R$ is a constant and $P$, $V$, and $T$ are the pressure, volume per mole, and temperature, respectively, of the gas. It will be proved that

$$\frac{dP}{P} + \frac{dV}{V} = \frac{dT}{T}.$$  \hspace{1cm} (3.3)

Recall that $dP$, $dV$, and $dT$ represent infitesimal changes in the quantities $P$, $V$, and $T$, respectively. Notice that none of the quotients in Equation (3.3) have an infinitesimal in the denominator. For example, $dP$ is divided not by $dx$ or $dt$, as it would be in a derivative such as $\frac{dP}{dx}$ or $\frac{dP}{dt}$. Instead it is divided by $P$, which is not an infinitesimal. So Equation (3.3) is an equation that relates infitesimals themselves, i.e. infinitesimal changes, not infinitesimal rates of change. This is, in fact, how many physical laws are stated, for reasons that will be discussed shortly.

Though infinitesimals have been used throughout this text, many calculus textbooks\(^7\) do not even mention them, instead preferring to call them *differentials*.\(^8\) For compatibility, the definition is given here:

For a differentiable function $f(x)$, the **differential** of $f(x)$ is

$$df = f'(x) \, dx$$  \hspace{1cm} (3.4)

where $dx$ is an infinitesimal change in $x$.

Note that this is identical to Equation (1.9) in Section 1.3.

**Example 3.32**

Find the differential $df$ of $f(x) = x^3$.

*Solution:* By definition,

$$df = f'(x) \, dx = 3x^2 \, dx$$

Equivalently, this can be written as

$$d(x^3) = 3x^2 \, dx,$$

which is often the way it would appear in textbooks in the sciences.

All the rules for derivatives (e.g. sum rule, product rule) apply to differentials, and can be proved simply by multiplying the corresponding derivative rule by $dx$ on both sides of the equation:

---

\(^7\)More accurately, many current calculus textbooks never mention them. Calculus texts up through the 1930s or so not only mentioned infinitesimals but used them extensively, even to the point of the texts themselves having titles such as *Introduction to Infinitesimal Calculus*.

\(^8\)Though often in an unclear and sometimes confusing and misleading manner, as will be seen later in this section.
Let $f$ and $g$ be differentiable functions, and let $c$ be a constant. Then:

(a) $d(c) = 0$

(b) $d(cf) = c \, df$ (Constant Multiple Rule)

(c) $d(f + g) = df + dg$ (Sum Rule)

(d) $d(f - g) = df - dg$ (Difference Rule)

(e) $d(fg) = f \, dg + g \, df$ (Product Rule)

(f) $d\left(\frac{f}{g}\right) = \frac{g \, df - f \, dg}{g^2}$ (Quotient Rule)

(g) $d(f^n) = nf^{n-1} \, df$ (Power Rule)

(h) $d(f(g)) = \frac{df}{dg} \, dg$ (Chain Rule)

For example, to prove (e), multiply both sides of the usual Product Rule by $dx$ so that

$$d\left(\frac{f}{g}\right) dx = f \frac{dg}{dx} + g \frac{df}{dx} \Rightarrow d\left(\frac{f}{g}\right) = dx \left(\frac{f \, dg}{dx} + g \frac{df}{dx}\right) \Rightarrow d\left(\frac{f}{g}\right) = f \, dg + g \, df$$

since the $dx$ terms all cancel. The proofs of the other rules are similar.

The differential version of the ideal gas law in Equation (3.3)

$$\frac{dP}{P} + \frac{dV}{V} = \frac{dT}{T}$$

can now be proved by taking the differential of both sides of the equation $PV = RT$:

$$d(PV) = d(RT) = R \cdot dT \text{ by the Constant Multiple Rule}$$

$$V \, dP + P \, dV = \frac{PV}{T} dT \text{ by the Product Rule and since } R = \frac{PV}{T}$$

$$\frac{V \, dP}{PV} + \frac{P \, dV}{PV} = \frac{dT}{T} \text{ after dividing both sides by } PV$$

$$\frac{dP}{P} + \frac{dV}{V} = \frac{dT}{T} \checkmark$$

Notice that $\frac{dP}{P}$, $\frac{dV}{V}$ and $\frac{dT}{T}$ represent the relative infinitesimal changes in $P$, $V$, and $T$, respectively. The differential formulation is useful for finding one relative infinitesimal change when the other two are known.
Example 3.33

Suppose that $M$ is the total mass of a rocket and its unburnt fuel at any time $t$ (so $M$ is a function of $t$). Over an infinitesimal time $dt$ a mass $dm$ of fuel is burnt and the gas byproducts are expelled out the rear of the rocket at a velocity $v_E$ relative to the rocket. Using the law of conservation of momentum over the interval $dt$, show that

$$v_E dm = M dv$$

where $m$ and $v$ are the mass of burnt fuel and the velocity of the rocket, respectively, at the beginning of the time $dt$.

**Solution:** Momentum is defined as mass times velocity. The momentum of the rocket at the beginning of the time $dt$ is thus $Mv$. At the end of the time $dt$, the momentum of the rocket consists of two parts, namely the momentum of the rocket and its remaining unburnt fuel, which is

$$(\text{mass before } dt) - (\text{increase in burnt fuel}) \times (\text{(velocity before } dt) + (\text{increase in velocity}))$$

and the momentum of the fuel that was burnt and expelled out the rear, which is

$$(v - v_E)dm.$$

So by conservation of momentum,

$$Mv = (M - dm)(v + dv) + (v - v_E)dm$$

$$Mv = Mv - v dm + M dv - (dm)(dv) + v dm - v_E dm, \quad \text{so}$$

$$v_E dm = M dv - (dm)(dv) = M dv$$

since $(dm)(dv) = (m'(t)dt)(v'(t)dt) = m'(t)v'(t)(dt)^2 = m'(t)v'(t) \cdot 0 = 0$.

Dividing both sides of $v_E dm = M dv$ by $dt$ yields the equation

$$M \dot{v} = \dot{m} v_E$$

using the dot notation—mentioned in Section 1.3—for the derivative with respect to the time variable $t$, which is still popular with physicists. Since $\dot{v}$ is just acceleration $a$, this formulation is the classic equation for the acceleration of a rocket.\(^9\)

Letting $f$ be the natural logarithm function and letting $g = u$ in the differential version of the Chain Rule yields the following useful result:

$$d(\ln u) = \frac{du}{u}$$

This is often used in a differential version of the technique of logarithmic differentiation discussed in Section 2.3.

Example 3.34

Prove the relation \( \frac{dP}{P} + \frac{dV}{V} = \frac{dT}{T} \) using logarithmic differentiation.

**Solution:** Take the natural logarithm and then the differential of both sides of the equation \( PV = RT \):

\[
\ln(PV) = \ln(RT) \quad \Rightarrow \quad \ln P + \ln V = \ln R + \ln T \\
\Rightarrow \quad d(\ln P + \ln V) = d(\ln R + \ln T) \\
\Rightarrow \quad \frac{dP}{P} + \frac{dV}{V} = 0 + \frac{dT}{T} = \frac{dT}{T} \quad \text{(since \( \ln R \) is a constant)}
\]

Example 3.35

The derivative of the area \( \pi r^2 \) of a circle of radius \( r \), as a function of \( r \), equals its circumference \( 2\pi r \). Use the notion of a differential as an infinitesimal change to explain why this makes sense geometrically.

**Solution:** Let \( A = \pi r^2 \) be the area of a circle of varying radius \( r \). Then \( A'(r) = 2\pi r \), which is equivalent to saying \( dA = 2\pi r \, dr \). To see why this makes sense geometrically, imagine increasing the radius by \( dr \), as in the picture below on the left. This increases the area \( A \) of the circle to \( A + dA \), with \( dA \) the infinitesimal area of the shaded ring in the picture.

[Diagram of a circle with \( dr \) and \( dA \) labeled]

Slice that ring along the dashed line then roll it flat, yielding a trapezoid with height \( dr \), top length \( 2\pi r \) (from the circumference of the inner circle of the ring), and bottom length \( 2\pi(r + dr) \) (from the circumference of the outer circle of the ring), as shown in the picture above on the right. The triangular edges of the trapezoid contribute nothing to the area of the trapezoid, since (by the Microstraightness Property) the hypotenuse of each is indeed a straight line, so each is a right triangle with height \( dr \) and (by symmetry) base \( \pi dr \), thus having area \( \frac{1}{2} \pi (dr)^2 = 0 \). Hence the entire area \( dA \) of the trapezoid comes from the rectangular portion of height \( dr \) and base \( 2\pi r \), which means \( dA = 2\pi r \, dr \), as expected.

The above example answers the question of whether it is a happy coincidence that the derivative of a circle’s area turns out to be the circle’s circumference—no, it is not! Some other such cases (e.g. the derivative of a sphere’s volume is its surface area) are left to the exercises. Note that a similar “coincidence” does not occur for a square: if \( x \) is the length of each side then the area is \( x^2 \), but the derivative of \( x^2 \) is \( 2x \), which is not the perimeter of the square (i.e. \( 4x \)). Why does this not follow the same pattern as the circle? Think about a key difference in the shape of a square in comparison to a circle, keeping differentiability in mind.
There are many benefits to using differentials—i.e. infinitesimals—in calculus.\footnote{For an excellent overview on this subject, see DRAY, T. AND C.A. MANOGUE, Putting Differentials Back into Calculus, College Math. J. 41 (2010), 90-100. Some of the material in this section is indebted to that paper, which is available at http://www.math.oregonstate.edu/bridge/papers/differentials.pdf} For example, recall Example 3.31 in Section 3.5 on related rates, where a right circular cylinder’s volume $V$—with radius $r$ and height $h$—changes with time $t$ as

$$\frac{dV}{dt} = (2\pi r \cdot \frac{dr}{dt}) h + \pi r^2 \cdot \frac{dh}{dt}.$$ 

The above equation forces you to consider only the derivative with respect to the time variable $t$. What if you wanted to see the rates of change with respect to another variable, such as $r$, $h$, or some other quantity? In that case using the differential version of the above equation, namely

$$dV = 2\pi rh \, dr + \pi r^2 \, dh$$

provides more flexibility—you are free to divide both sides by any differential, not just by $dt$. Many related rates problems would likely benefit from this approach.

Present-day calculus textbooks confuse the notion of a differential (infinitesimal) $dx$ with the idea of a small but real value $\Delta x$. The two are not the same. An infinitesimal is not a real number and cannot be assigned a real value, no matter how small; $\Delta x$ can be assigned real values. Using $dx$ and $\Delta x$ interchangeably is a source of much confusion for students (likewise for $dy$ and $\Delta y$). This confusion rears its head in exercises involving the linear approximation of a curve by its tangent line near a point $x_0$, namely $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$ when $x - x_0$ is “small” (e.g. $\sqrt{63} \approx 7.9375$, by using $f(x) = \sqrt{x}$, $x = 63$, $x_0 = 64$, and $x - x_0 = \Delta x = -1$). Such exercises have nothing to do with differentials, not to mention being of dubious value nowadays (people, really). They are remnants of a bygone era, before the advent of modern computing that obviates the need for such (generally) poor approximations.

### Exercises

A

1. Find the differential $df$ of $f(x) = x^2 - 2x + 5$.
2. Find the differential $df$ of $f(x) = \sin^2(x^2)$.
3. Show that $d(\tan^{-1}(y/x)) = \frac{x \, dy - y \, dx}{x^2 + y^2}$.
4. Given $y^2 - xy + 2x^2 = 3$, find $dy$.
5. The elasticity of a function $y = f(x)$ is $E(y) = \frac{x \, dy}{y \, dx}$. Show that $E(y) = \frac{d(\ln y)}{d(\ln x)}$.
6. Prove the differential version of the Quotient Rule:

$$d\left(\frac{f}{g}\right) = \frac{g \, df - f \, dg}{g^2}$$
7. Let \( y = c u^n \), where \( c \) and \( n \) are constants. Show that
\[
\frac{dy}{y} = n \frac{du}{u}.
\]

8. Obviously the derivative of \( \pi^2 \) is not \( 2\pi \). But is \( d(\pi^2) = 2\pi d(\pi) \) true? Explain.

9. The continuity relation for an ideal gas is
\[
\frac{PM}{\sqrt{T}} = \text{constant}
\]
where \( P \) and \( T \) are the pressure and temperature, respectively, of the gas, and \( M \) is the Mach number. Show that
\[
\frac{dP}{P} + \frac{dM}{M} = \frac{dT}{2T}.
\]

10. For an ideal gas, satisfying the equation \( PV = RT \) as before, the Gibbs energy \( G \) of an ideal gas is defined as \( G = H - TS \), where \( H \) and \( S \) are the enthalpy and entropy, respectively, of the gas.
   (a) Show that
   \[
   d\left( \frac{G}{RT} \right) = \frac{1}{RT} dG - \frac{G}{R T^2} dT.
   \]
   (b) One of the fundamental property relations for an ideal gas (which you do not need to prove) is
   \[
   dG = V \, dP - S \, dT.
   \]
   Use this and part(a) to show that
   \[
   d\left( \frac{G}{RT} \right) = \frac{V}{RT} dP - \frac{H}{R T^2} dT.
   \]

11. The derivative of the volume \( \pi r^2 h \) of a right circular cylinder of radius \( r \) and height \( h \), as a function of \( r \), equals its lateral surface area \( 2\pi rh \). Use the notion of a differential as an infinitesimal change to explain why this makes sense geometrically.

12. The derivative of the volume \( \frac{4}{3} \pi r^3 \) of a sphere of radius \( r \), as a function of \( r \), equals its surface area \( 4\pi r^2 \). Use the notion of a differential as an infinitesimal change to explain why this makes sense geometrically.

13. In quantum calculus the \( q \)-differential of a function \( f(x) \) is
\[
d_q f(x) = f(qx) - f(x),
\]
and the \( q \)-derivative of \( f(x) \) is
\[
D_q f(x) = \frac{d_q f(x)}{d_q x} = \frac{f(qx) - f(x)}{q^x - x} = \frac{f(qx) - f(x)}{(q-1)x}.
\]
   (a) Show that for all positive integers \( n \),
   \[
   D_q (x^n) = [n]_q x^{n-1},
   \]
   where \([n]_q = 1 + q + q^2 + \cdots + q^{n-1}\).
   (b) Use part(a) to show that for all positive integers \( n \),
   \[
   \lim_{q \to 1} D_q (x^n) = \frac{d}{dx} (x^n).
   \]
CHAPTER 4

Applications of Derivatives

4.1 Optimization

Many physical problems involve optimization: finding either a maximum or minimum value of some quantity. For example, suppose you want to throw a rock as far as possible. If you always throw with the same speed, at what angle with the ground should the rock be thrown to maximize the horizontal distance traveled by the rock? This section will present some techniques for solving such problems.

First, the intuitive notions of maximum and minimum need clarifying.

A function $f$ has a **global maximum** at $x = c$ if $f(c) \geq f(x)$ for all $x$ in the domain of $f$. Similarly, $f$ has a **global minimum** at $x = c$ if $f(c) \leq f(x)$ for all $x$ in the domain of $f$. Say that $f$ has a **local maximum** at $x = c$ if $f(c) \geq f(x)$ for all $x$ “near” $c$, i.e. for all $x$ such that $|x - c| < \delta$ for some number $\delta > 0$. Likewise, $f$ has a **local minimum** at $x = c$ if $f(c) \leq f(x)$ for all $x$ such that $|x - c| < \delta$ for some number $\delta > 0$.

In other words, a global maximum is the largest value everywhere (“globally”), whereas a local maximum is only the largest value “locally.” Likewise for a global vs local minimum. The picture below illustrates the differences.

In the picture, on the interval $[a, b]$ the function $f$ has a global minimum at $x = a$, a global maximum at $x = c_1$, a local minimum at $x = c_2$, and a local maximum at $x = b$. 


Every global maximum [minimum] is a local maximum [minimum], but not vice versa. In physical applications global maxima or minima\footnote{The words “maxima” and “minima” are the traditional plural forms of maximum and minimum, respectively.} are the primary interest. The Extreme Value Theorem in Section 3.3 guarantees the existence of at least one global maximum and at least one global minimum for continuous functions defined on closed intervals (i.e. intervals of the form \([a, b]\)). All the functions under consideration here will be differentiable, and hence continuous. So the only issues will be how to find the global maxima or minima, and how to handle intervals that are not closed.

Consider again the picture from the previous page, this time looking at how the derivative \(f'\) changes over \([a, b]\). Intuitively it is obvious that near an internal maximum (i.e. in the open interval \((a, b)\)) such as at \(x = c_1\), the function should increase before that point and then decrease after that point. That means that \(f'(x) > 0\) before \(x = c_1\) and \(f'(x) < 0\) after the “turning point” \(x = c_1\), as shown below.

\[
\begin{array}{c}
\text{Assuming that } f' \text{ is continuous (which will be the case for all the functions in this section), then this means that } f'(c_1) = 0. \text{ Similarly, near the internal minimum at } x = c_2, \text{ } f'(x) < 0 \text{ before } x = c_2 \text{ and } f'(x) > 0 \text{ after } x = c_2, \text{ so that } f''(c_2) = 0. \text{ Points at which the derivative is zero are called critical points (or stationary points) of the function. So } x = c_1 \text{ and } x = c_2 \text{ are critical points of } f. \\
\text{Note in the picture that } f' \text{ goes from positive to zero to negative around } x = c_1, \text{ so that } f' \text{ is decreasing around } x = c_1, \text{ i.e. } f''(c_1) < 0. \text{ Similarly, } f' \text{ is increasing around } x = c_2, \text{ i.e. } f''(c_2) > 0. \text{ This leads to the following test for local maxima and minima: }^2
\end{array}
\]

\[\text{Second Derivative Test: Let } x = c \text{ be a critical point of } f \text{ (i.e } f'(c) = 0) \text{. Then:}
\]

(a) If \(f''(c) > 0\) then \(f\) has a local minimum at \(x = c\).

(b) If \(f''(c) < 0\) then \(f\) has a local maximum at \(x = c\).

(c) If \(f''(c) = 0\) then the test fails.

To see why the test fails when \(f''(c) = 0\), consider \(f(x) = x^2\): \(f'(0) = 0\) and \(f''(0) = 0\), yet \(x = 0\) is neither a local minimum nor maximum in any open interval containing \(x = 0\). Section 4.2 will present an alternative for when the Second Derivative Test fails.

\[\footnote{A formal proof requires the Mean Value Theorem, which will be presented in Section 4.4.}\]
There is a simple visual mnemonic device for remembering the Second Derivative Test, due to a generic minimum or maximum resembling a smile or frown, respectively:

- $f'' > 0$ local min.
- $f'' < 0$ local max.
- $f'' = 0$ test fails

The “eyes” in the faces represent the sign of $f''$ at a critical point, while the “mouths” indicate the nature of that point (when $f'' = 0$ nothing is known). The procedure for finding a global maximum or minimum can now be stated:

**How to find a global maximum or minimum**

Suppose that $f$ is defined on an interval $I$. There are two cases:

1. **The interval $I$ is closed**: The global maximum of $f$ will occur either at an interior local maximum or at one of the endpoints of $I$ whichever of these points provides the largest value of $f$ will be where the global maximum occurs. Similarly, the global minimum of $f$ will occur either at an interior local minimum or at one of the endpoints of $I$; whichever of these points provides the smallest value of $f$ will be where the global minimum occurs.

2. **The interval $I$ is not closed and has only one critical point**: If the only critical point is a local maximum then it is a global maximum. If the only critical point is a local minimum then it is a global minimum.

In each case of the above procedure try to use the Second Derivative Test to verify that a critical point is a local minimum or maximum, unless it is obvious from the nature of the problem that there can be only a minimum or only a maximum.

**Example 4.1**

Show that the rectangle with the largest area for a fixed perimeter is a square.

**Solution**: Let $L$ be the perimeter of a rectangle with sides $x$ and $y$. The idea is that $L$ is a fixed constant, but $x$ and $y$ can vary. Figure 4.1.1 shows that there are many possible shapes for the rectangle, but in all cases $L = 2x + 2y$. Let $A$ be the area of such a rectangle. Then $A = xy$, which is a function of two variables. But

$$L = 2x + 2y \Rightarrow y = \frac{L}{2} - x ,$$

![Figure 4.1.1](image-url)
and hence

\[ A = x \left( \frac{L}{2} - x \right) = \frac{Lx}{2} - x^2 \]

is now a function of \( x \) alone, on the open interval \((0, L/2)\) (since the length \( x \) is positive). Now find the critical points of \( A \):

\[ A'(x) = 0 \Rightarrow \frac{L}{2} - 2x = 0 \]
\[ \Rightarrow x = \frac{L}{4} \text{ is the only critical point} \]

This problem is thus the case of a function defined on an open interval having only one critical point. Use the Second Derivative Test to verify that the sole critical point \( x = L/4 \) is a local maximum for \( A \):

\[ A''(x) = -2 \Rightarrow A''(L/4) = -2 < 0 \Rightarrow A \text{ has a local maximum at } x = L/4 \]

Thus, \( A \) has a global maximum at \( x = L/4 \). Also, \( y = L/2 - x = L/2 - L/4 = L/4 \), which means that \( x = y \), i.e. the rectangle is a square.

Example 4.2

Suppose a right circular cylindrical can with top and bottom lids will be assembled to have a fixed volume. Find the radius and height of the can that minimizes the total surface area of the can.

Solution: Let \( V \) be the fixed volume of the can with radius \( r \) and height \( h \), as in Figure 4.1.2. The volume \( V \) is a constant, with \( V = \pi r^2 h \). Let \( S \) be the total surface area of the can, including the lids. Then

\[ S = 2\pi r^2 + 2\pi rh \]

where the first term in the sum on the right side of the equation is the combined area of the two circular lids and the second term is the lateral surface area of the can. So \( S \) is a function of \( r \) and \( h \), but \( h \) can be eliminated since

\[ V = \pi r^2 h \Rightarrow h = \frac{V}{\pi r^2} \]

and so

\[ S = 2\pi r^2 + 2\pi r \cdot \frac{V}{\pi r^2} = 2\pi r^2 + \frac{2V}{r} \]

making \( S \) a function of \( r \) alone. Now find the critical points of \( S \) (i.e. solve \( S'(r) = 0 \)):

\[ S'(r) = 0 \Rightarrow 4\pi r - \frac{2V}{r^2} = 0 \]
\[ \Rightarrow r^3 = \frac{V}{2\pi} \]
\[ \Rightarrow r = \sqrt[3]{\frac{V}{2\pi}} \text{ is the only critical point} \]
Since both \( r \) and \( h \) are lengths and have to be positive, then \( 0 < r < \infty \). So this is another case of a function defined on an open interval having only one critical point. Use the Second Derivative Test to verify that this critical point \( r = \sqrt[3]{\frac{V}{2\pi}} \) is a local minimum for \( S \):

\[
S''(r) = 4\pi + \frac{4V}{r^3} \quad \Rightarrow \quad S''\left(\sqrt[3]{\frac{V}{2\pi}}\right) = 4\pi + \frac{4V}{\sqrt[3]{V^2}} = 12\pi > 0 \quad \Rightarrow \quad S \text{ has a local minimum at } r = \sqrt[3]{\frac{V}{2\pi}}
\]

Thus, \( S \) has a global minimum at \( r = \sqrt[3]{\frac{V}{2\pi}} \), and

\[
r \cdot r^2 = r^3 = \frac{V}{2\pi} \quad \Rightarrow \quad 2r = \frac{V}{\pi r^2} = h.
\]

Hence, \( r = \sqrt[3]{\frac{V}{2\pi}} \) and \( h = 2\sqrt[3]{\frac{V}{2\pi}} \) will minimize the total surface area, i.e. the height should equal the diameter.

Note that this result can be applied to soda cans, where the volume is \( V = 12 \) fluid ounces \( \approx 21.6 \) cubic inches: both a diameter and height of about 3.8 inches will minimize the amount (and hence the cost) of the aluminum used for the can. Yet soda cans are not that wide and short—they are usually thinner and taller. So why is a non-optimal size used in practice? Other factors—e.g. packing requirements, the need for small children to hold the can in one hand—might override the desire to minimize the cost of the aluminum. The lesson is that an optimal solution for one factor (material cost) might not always be truly optimal when all factors are considered; compromise is often necessary.

**Example 4.3**

Suppose that a projectile is launched from the ground with a fixed initial velocity \( v_0 \) at an angle \( \theta \) with the ground. What value of \( \theta \) would maximize the horizontal distance traveled by the projectile, assuming the ground is flat and not sloped (i.e. horizontal)?

**Solution:** Let \( x \) and \( y \) represent the horizontal position and vertical position, respectively, of the projectile at time \( t \geq 0 \). From the triangle at the bottom of Figure 4.1.3, the horizontal and vertical components of the initial velocity are \( v_0 \cos \theta \) and \( v_0 \sin \theta \), respectively. Since distance is the product of velocity and time, then the horizontal and vertical distances traveled by the projectile by time \( t \) due to the initial velocity are \((v_0 \cos \theta) t \) and \((v_0 \sin \theta) t \), respectively. Ignoring wind and air resistance, the only other force on the projectile will be the downward force \( g \) due to gravity, so that the equations of motion for the projectile are:

\[
x = (v_0 \cos \theta) t \quad \Rightarrow \quad t = \frac{x}{v_0 \cos \theta} \quad \Rightarrow \quad y = -\frac{1}{2} g t^2 + (v_0 \sin \theta) t
\]

The goal is to find \( \theta \) that maximizes the length \( L \) shown in Figure 4.1.3. First write \( y \) as a function of \( x \):

\[
x = (v_0 \cos \theta) t \quad \Rightarrow \quad t = \frac{x}{v_0 \cos \theta} \quad \Rightarrow \quad y = -\frac{1}{2} g \left( \frac{x}{v_0 \cos \theta} \right)^2 + (v_0 \sin \theta) \cdot \frac{x}{v_0 \cos \theta}
\]

\[
\Rightarrow y = -\frac{gx^2}{2v_0^2 \cos^2 \theta} + x \tan \theta
\]
Then $L$ is the value of $x > 0$ that makes $y = 0$:

$$0 = -\frac{gL^2}{2v_0^2 \cos^2 \theta} + L \tan \theta \quad \Rightarrow \quad L = \frac{2v_0^2 \sin \theta \cos \theta}{g} = \frac{v_0^2 \sin 2\theta}{g}$$

So $L$ is now a function of $\theta$, with $0 < \theta < \pi/2$ (why?). So if there is a single local maximum then it must be the global maximum. Now get the critical points of $L$:

$$L'(\theta) = 0 \quad \Rightarrow \quad \frac{2v_0^2 \cos 2\theta}{g} = 0$$

$$\Rightarrow \quad \cos 2\theta = 0$$

$$\Rightarrow \quad 2\theta = \frac{\pi}{2} \quad \Rightarrow \quad \theta = \frac{\pi}{4}$$

is the only critical point

Use the Second Derivative Test to verify that $L$ has a local maximum at $\theta = \pi/4$:

$$L''(\theta) = -\frac{4v_0^2 \sin 2\theta}{g} \quad \Rightarrow \quad L''(\pi/4) = -\frac{4v_0^2}{g} < 0$$

$$\Rightarrow \quad L$$ has a local maximum at $\theta = \frac{\pi}{4}$.

Thus, $L$ has a global maximum at $\theta = \frac{\pi}{4}$, i.e. the projectile travels the farthest horizontally when launched at a $45^\circ$ angle with the ground.

Note that once the formula for $L$ as a function of $\theta$ was found to be $L = \frac{v_0^2 \sin 2\theta}{g}$, calculus was not actually needed to solve this problem. Why? Since $v_0^2$ and $g$ are positive constants (recall $g = 9.8\text{m/s}^2$), $L$ would have its largest value when $\sin 2\theta$ has its largest value 1, which occurs when $\theta = \pi/4$.

**Example 4.4**

Fermat’s Principle states that light always travels along the path that takes the least amount of time. So suppose that a ray of light is shone from a point $A$ onto a flat horizontal reflective surface at an angle $\theta_1$ with the surface and then reflects off the surface at an angle $\theta_2$ to a point $B$. Show that Fermat’s Principle implies that $\theta_1 = \theta_2$.

**Solution:** Let $L$ be the horizontal distance between $A$ and $B$, let $d_1$ be the distance the light travels from $A$ to the point of contact $C$ with the surface a horizontal distance $x$ from $A$, let $d_2$ be the distance from $C$ to $B$, and let $y_1$ and $y_2$ be the vertical distances from $A$ and $B$, respectively, to the surface, as in the picture below.
Since time is distance divided by speed, and since the speed of light is constant, then minimizing the total time elapsed is equivalent to minimizing the total distance traveled, namely \( D = d_1 + d_2 \). The basic idea here is that Fermat’s Principle implies that for the light to go from \( A \) to \( B \) in the shortest time, the unknown point \( C \)—and hence the unknown distance \( x \)—will have to be at a point that makes \( \theta_1 = \theta_2 \). The distances \( L, y_1 \) and \( y_2 \) are constants, so the goal is to write the total distance \( D \) as a function of \( x \), find the \( x \) that minimizes \( D \), then show that that value of \( x \) makes \( \theta_1 = \theta_2 \).

First, note that \( C \) has to be between \( A \) and \( B \) as in the picture, otherwise the total distance \( D \) would be larger than if \( C \) were directly below either \( A \) or \( B \). This ensures that \( \theta_1 \) and \( \theta_2 \) are between 0 and \( \pi/2 \), and that \( 0 \leq x \leq L \).

Next, by the Pythagorean Theorem and the above picture,

\[
d_1 = \sqrt{x^2 + y_1^2} \quad \text{and} \quad d_2 = \sqrt{(L-x)^2 + y_2^2}
\]

and so the total distance \( D = d_1 + d_2 \) traveled by the light is a function of \( x \):

\[
D(x) = \sqrt{x^2 + y_1^2} + \sqrt{(L-x)^2 + y_2^2}
\]

To find the critical points of \( D \), solve the equation \( D'(x) = 0 \):

\[
D'(x) = \frac{x}{\sqrt{x^2 + y_1^2}} - \frac{L-x}{\sqrt{(L-x)^2 + y_2^2}} = 0 \quad \Rightarrow \quad \frac{x}{d_1} = \frac{L-x}{d_2} \quad \Rightarrow \quad \sin \theta_1 = \sin \theta_2 \quad \Rightarrow \quad \theta_1 = \theta_2
\]

since the sine function is one-to-one over the interval \([0, \pi/2] \).

This seems to prove the result, except for one remaining issue to resolve: verifying that the minimum for \( D \) really does occur at the \( x \) between 0 and \( L \) where \( D'(x) = 0 \), not at the endpoints \( x = 0 \) or \( x = L \) of the closed interval \([0,L] \). Note that using the Second Derivative Test in this case does not matter, since you would have to check the value of \( D \) at the endpoints anyway and compare those values to the value of \( D \) at the critical point. To find an expression for critical point, note that

\[
D'(x) = 0 \quad \Rightarrow \quad \frac{x}{\sqrt{x^2 + y_1^2}} = \frac{L-x}{\sqrt{(L-x)^2 + y_2^2}} \quad \Rightarrow \quad \frac{x^2}{x^2 + y_1^2} = \frac{(L-x)^2}{(L-x)^2 + y_2^2}
\]

\[
= (L-x)^2 \frac{x}{x^2 + y_1^2} + x^2 \frac{y_2}{x^2 + y_1^2} = (L-x)^2 \frac{x}{x^2 + y_1^2} + (L-x)^2 \frac{y_2}{x^2 + y_1^2}
\]

\[
x y_2 = (L-x)y_1 \quad \Rightarrow \quad x = \frac{L y_1}{y_1 + y_2} \quad \text{is the only critical point},
\]

and \( x \) is between 0 and \( L \). Now compare the values of \( D^2(x) \) at \( x = 0, x = L, \) and \( x = \frac{L y_1}{y_1 + y_2} \):

\[
D^2(0) = L^2 + y_1^2 + y_2^2 + 2 y_1 \sqrt{L^2 + y_2^2}
\]

\[
D^2(L) = L^2 + y_1^2 + y_2^2 + 2 y_2 \sqrt{L^2 + y_1^2}
\]

\[
D^2\left(\frac{L y_1}{y_1 + y_2}\right) = L^2 + y_1^2 + y_2^2 + 2 y_1 y_2
\]

Since \( y_2 < \sqrt{L^2 + y_1^2} \) and \( y_1 < \sqrt{L^2 + y_1^2} \), then \( D^2\left(\frac{L y_1}{y_1 + y_2}\right) \) is the smallest of the three values above, so that \( D^2(x) \) has its minimum value at \( x = \frac{L y_1}{y_1 + y_2} \), which means \( D(x) \) has its minimum value there. \( \checkmark \)
Example 4.5

A man is in a boat 4 miles off a straight coast. He wants to reach a point 10 miles down the coast in the minimum possible time. If he can row 4 mi/hr and run 5 mi/hr, where should he land the boat?

Solution: Let \( T \) be the total time traveled. The goal is to minimize \( T \). From the picture on the right, since time is distance divided by speed, break the total time into two parts: the time rowing in the water and the time running on the coast, so that

\[
T = \text{time}_{\text{row}} + \text{time}_{\text{run}} = \frac{\text{dist}_{\text{row}}}{\text{speed}_{\text{row}}} + \frac{\text{dist}_{\text{run}}}{\text{speed}_{\text{run}}} = \frac{\sqrt{x^2 + 16}}{4} + \frac{10 - x}{5},
\]

where \( 0 \leq x \leq 10 \) is the distance along the coast where the boat lands. Then

\[
T'(x) = \frac{x}{4 \sqrt{x^2 + 16}} - \frac{1}{5} = 0 \quad \Rightarrow \quad 5x = 4 \sqrt{x^2 + 16} \quad \Rightarrow \quad 25x^2 = 16(x^2 + 16) \quad \Rightarrow \quad x = \frac{16}{3},
\]

so \( x = \frac{16}{3} \) is the only critical point. Thus, the (global) minimum of \( T \) will occur at \( x = \frac{16}{3}, x = 0 \), or \( x = 10 \). Since

\[
T(0) = 3, \quad T(10) = \frac{\sqrt{29}}{2} \approx 2.693, \quad T\left(\frac{16}{3}\right) = \frac{13}{5} = 2.6
\]

then \( T\left(\frac{16}{3}\right) < T(10) < T(0) \). Hence, the global minimum occurs when landing the boat \( x = \frac{16}{3} \approx 5.33 \) miles down the coast.

Note that this example shows the importance of checking the endpoints. It was very close between landing about 5.33 miles down the coast (2.6 hours) or simply rowing all the way to the destination (about 2.693 hours)—the difference is only about 5.6 minutes. With just a slight change in a few of the numbers, the minimum could have occurred at an endpoint. Moral: always check the endpoints\(^3\)

Example 4.6

Find the point \((x, y)\) on the graph of the curve \( y = \sqrt{x} \) that is closest to the point \((1, 0)\).

Solution: Let \((x, y)\) be a point on the curve \( y = \sqrt{x} \). Then \((x, y) = (x, \sqrt{x})\), so by the distance formula the distance \( D \) between \((x, y)\) and \((1, 0)\) is given by

\[
D^2 = (x - 1)^2 + (y - 0)^2 = (x - 1)^2 + (\sqrt{x})^2 = (x - 1)^2 + x,
\]

which is a function of \( x \geq 0 \) alone. Note that minimizing \( D \) is equivalent to minimizing \( D^2 \). Since

\[
\frac{d(D^2)}{dx} = 2(x - 1) + 1 = 2x - 1 = 0 \quad \Rightarrow \quad x = \frac{1}{2}
\]

is the only critical point, and since \( \frac{d^2(D^2)}{dx^2} = 2 > 0 \) for all \( x \), then by the Second Derivative Test \( x = 1/2 \) is a local minimum. Hence, the global minimum for \( D^2 \) must occur at the endpoint \( x = 0 \) or at \( x = 1/2 \). But \( D^2(0) = 1 > D^2(1/2) = 3/4 \), so the global minimum occurs at \( x = 1/2 \). Hence, the closest point is \((x, y) = (1/2, \sqrt{1/2})\).

\(^3\)Another possible lesson is that optimal in the mathematical sense might, again, not mean optimal in a practical sense. After all, presumably after the man is finished with whatever he had to do at the destination 10 miles down the coast, he then has the inconvenience of going back about 4.67 miles to retrieve his boat. At his running speed of 10 mph this would take 28 minutes, wiping out the 5.6 minutes he gained with his “optimal” landing spot!
Example 4.7

Find the width and height of the rectangle with the largest possible perimeter inscribed in a semicircle of radius $r$.

Solution: Let $w$ be the width of the rectangle and let $h$ be the height, as in the picture. Then the perimeter is $P = 2w + 2h$. By symmetry and the Pythagorean Theorem,

$$h^2 = r^2 - \left(\frac{w}{2}\right)^2 \Rightarrow h = \frac{1}{2} \sqrt{4r^2 - w^2}$$

and so $P = 2w + \sqrt{4r^2 - w^2}$ for $0 < w < 2r$. Find the critical points of $P$:

$$P'(w) = 2 - \frac{w}{\sqrt{4r^2 - w^2}} = 0 \Rightarrow w = 2\sqrt{4r^2 - w^2}$$

$$\Rightarrow w^2 = 16r^2 - 4w^2$$

$$\Rightarrow w = \frac{4r}{\sqrt{5}}$$ is the only critical point,

and since

$$P''(w) = -\frac{4r^2}{(4r^2 - w^2)^{3/2}} \Rightarrow P''\left(\frac{4r}{\sqrt{5}}\right) = -\frac{5^{3/2}}{2r} < 0$$

then $P$ has a local maximum at $w = \frac{4r}{\sqrt{5}}$, by the Second Derivative Test. Since $P(w)$ is defined for $w$ in the open interval $(0, 2r)$, the local maximum is a global maximum. For the width $w = \frac{4r}{\sqrt{5}}$, the height is $h = \frac{r}{\sqrt{5}}$, which gives the dimensions for the maximum perimeter.

Note: If $w$ were extended to include the cases of “degenerate” rectangles of zero width or height, i.e. $w = 0$ or $w = 2r$, then the maximum perimeter would still occur at $w = \frac{4r}{\sqrt{5}}$, since $P\left(\frac{4r}{\sqrt{5}}\right) = \frac{10r}{\sqrt{5}} \approx 4.472r$ is larger than $P(0) = 2r$ and $P(2r) = 4r$.

In general, optimization problems usually involve some kind of constraint which allows you to rewrite the original function to optimize—typically involving two variables—as a function of a single variable. This is done by using the constraint to solve for one variable in terms of another and then substituting that expression into the function to optimize. For instance, in Example 4.1 the constraint was $L = 2x + 2y$, and in Example 4.1 the constraint was $V = \pi r^2 h$.

Finding critical points might not always be easy—solving the equation $f'(x) = 0$ could be impossible in a simple closed form. In that case some numerical approximation technique will likely be necessary (and in most “real world” cases usually is). Numerical methods for finding roots of functions—a critical point for $f$ is a root of $f'$—will be discussed in Section 4.3.

Another dilemma is if the Second Derivative Test fails. In that case some other method will be required, which will be discussed in the next section. For the exercises in this section, however, that issue will not arise.
Exercises

A

1. A farmer wishes to fence a field bordering a straight stream with 1000 yd of fencing material. It is not necessary to fence the side bordering the stream. What is the maximum area of a rectangular field that can be fenced in this way?

2. The power output $P$ of a battery is given by $P = VI - RI^2$, where $I$, $V$, and $R$ are the current, voltage, and resistance, respectively, of the battery. If $V$ and $R$ are constant, find the current $I$ that maximizes $P$.

3. An impulse turbine consists of a high speed jet of water striking circularly mounted blades. The power $P$ developed by the turbine is given by $P = VU(V - U)$, where $V$ is the speed of the jet and $U$ is the speed of the turbine. If the jet speed $V$ is held constant, find the turbine speed $U$ that maximizes the power.

4. A man is in a boat 5 miles off a straight coast. He wants to reach a point 15 miles down the coast in the minimum possible time. If he can row 6 mi/hr and run 10 mi/hr, where should he land the boat?

5. The current $I$ in a voltaic cell is

$$I = \frac{E}{R + r},$$

where $E$ is the electromotive force and $R$ and $r$ are the external and internal resistance, respectively. Both $E$ and $r$ are internal characteristics of the cell, and hence can be treated as constants. The power $P$ developed in the cell is $P = RI^2$. For which value of $R$ is the power $P$ maximized?

6. Find the point(s) on the ellipse $\frac{x^2}{25} + \frac{y^2}{16} = 1$ closest to $(-1, 0)$.

7. Find the maximum area of a rectangle that can be inscribed in an ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, where $a > 0$ and $b > 0$ are arbitrary constants. Your answer should be in terms of $a$ and $b$.

8. Find the radius and angle of the circular sector with the maximum area and a fixed perimeter $P$.

9. Find the point on the curve $y = x^2$ that is closest to the point $(4, -1/2)$.

B

10. In an electric circuit with a supplied voltage (emf) $E$, a resistor with resistance $r_0$, and an inductor with reactance $x_0$, suppose you want to add a second resistor. If $r$ represents the resistance of this second resistor then the power $P$ delivered to that resistor is given by

$$P = \frac{E^2r}{(r + r_0)^2 + x_0^2},$$

with $E$, $r_0$, and $x_0$ treated as constants. For which value of $r$ is the power $P$ maximized?

11. The stress $\tau$ in the $xy$-plane along a varying angle $\phi$ is given by

$$\tau = \tau(\phi) = \frac{\sigma_x - \sigma_y}{2} \sin 2\phi + \tau_{xy} \cos 2\phi,$$

where $\sigma_x$, $\sigma_y$, and $\tau_{xy}$ are stress components that can be treated as constants. Show that the maximum stress is

$$\tau = \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + 4\tau_{xy}^2}.$$

(Hint: Draw a right triangle with angle $2\phi$ after finding the critical point(s).)
12. A rectangular poster is to contain 50 square inches of printed material, with 4-inch top and bottom margins and 2-inch side margins. What dimensions for the poster would use the least paper?

13. Find the maximum volume of a right circular cylinder that can be inscribed in a sphere of radius 3.

14. A figure consists of a rectangle whose top side coincides with the diameter of a semicircle atop it. If the perimeter of the figure is 20 m, find the radius and height of the semicircle and rectangle, respectively, that maximizes the area inside the figure.

15. A thin steel pipe 25 ft long is carried down a narrow corridor 5.4 ft wide. At the end of the corridor is a right-angle turn into a wider corridor. How wide must this corridor be in order to get the pipe around the corner? You may assume that the width of the pipe can be ignored.

16. A rectangle is inscribed in a right triangle, with one corner of the rectangle at the right angle of the triangle. Show that the maximum area of the rectangle occurs when a corner of the rectangle is at the midpoint of the hypotenuse of the triangle. (Hint: Place the right angle of the triangle at the origin in the $xy$-plane.)

17. Find the relation between the radius and height of a cylindrical can with an open top that maximizes the volume of the can, given that the surface area of the can is always the same fixed amount.

18. An isosceles triangle is circumscribed about a circle of radius $r$. Find the height of the triangle that minimizes the perimeter of the triangle.

19. Suppose $N$ voltaic cells are arranged in $N/x$ rows in parallel, with each row consisting of $x$ cells in series, creating a current $I$ through an external resistance $R$. Each cell has internal resistance $r$ and EMF (voltage) $e$. Find the $x$ that maximizes the current $I$, which—due to Ohm’s Law—is given by

$$I = \frac{xe}{(x^2r/N) + R}.$$ 

20. Recall Fermat’s Principle from Example 4.4, which states that light travels along the path that takes the least amount of time. The speed of light in a vacuum is approximately $c = 2.998 \times 10^8$ m/s, but in some other medium (e.g. water) light is slower. Suppose that a ray of light goes from a point $A$ in one medium where it moves at a speed $v_1$ and ends up at a point $B$ in another medium where it moves at a speed $v_2$. Use Fermat’s Principle to prove Snell’s Law, which says that the light is refracted through the boundary between the two media such that

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2}$$

where $\theta_1$ and $\theta_2$ are the angles that the light makes with the normal line perpendicular to the boundary of the media in the first and second medium, respectively, as in the picture above.

21. A sphere of radius $a$ is inscribed in a right circular cone, with the sphere touching the base of the cone. Find the radius and height of the cone if its volume is a minimum.

22. Find the length of the shortest line segment from the positive $x$-axis to the positive $y$-axis going through a point $(a, b)$ in the first quadrant.

23. Find the radius $r$ of a circle $c$ whose center is on a fixed circle $C$ of radius $R$ such that the arc length of the part of $c$ within $C$ is a maximum.
4.2 Curve Sketching

A function can increase between two points in different ways, as shown in Figure 4.2.1.

\[ y = f(x) \]

(a) \( f'' = 0 \)

(b) \( f'' > 0: \text{concave up} \)

(c) \( f'' < 0: \text{concave down} \)

**Figure 4.2.1** Increasing function \( f: f' > 0 \), different signs for \( f'' \)

In each case in the above figure the function is increasing, so that \( f'(x) > 0 \), but the manner in which the function increases is determined by its **concavity**, that is, by the sign of the second derivative \( f''(x) \). The function in the graph on the far left is linear, i.e. of the form \( f(x) = ax + b \) for some constants \( a \) and \( b \), so that \( f''(x) = 0 \) for all \( x \). But the functions in the other two graphs are nonlinear. In the middle graph the derivative \( f' \) is increasing, so that \( f'' > 0 \); in this case the function is called **concave up**. In the graph on the far right the derivative \( f' \) is decreasing, so that \( f'' < 0 \); in this case the function is called **concave down**. The same definitions would hold if the function were decreasing, as shown in Figure 4.2.2 below:

\[ y = f(x) \]

(a) \( f'' = 0 \)

(b) \( f'' > 0: \text{concave up} \)

(c) \( f'' < 0: \text{concave down} \)

**Figure 4.2.2** Decreasing function \( f: f' < 0 \), different signs for \( f'' \)

In Figures 4.2.1(b) and 4.2.2(b) the function is below the line joining the points at each end, while in Figure 4.2.1(c) and 4.2.2(c) the function is above that line. This turns out to be true in general, as a result of the following theorem:
Concavity Theorem: Suppose that \( f \) is a twice-differentiable function on \([a, b]\). Then:

(a) If \( f''(x) > 0 \) on \((a, b)\) then \( f(x) \) is below the line \( l(x) \) joining the points \((a, f(a))\) and \((b, f(b))\) for all \( x \) in \((a, b)\).

(b) If \( f''(x) < 0 \) on \((a, b)\) then \( f(x) \) is above the line \( l(x) \) joining the points \((a, f(a))\) and \((b, f(b))\) for all \( x \) in \((a, b)\).

Proof: Only part (a) will be proved; the proof of part (b) is similar and left as an exercise. So assume that \( f''(x) > 0 \) on \((a, b)\), and \( l(x) \) be the line joining \((a, f(a))\) and \((b, f(b))\), as in the drawing on the right. The drawing suggests that \( f(x) < l(x) \) over \((a, b)\), but this is what has to be proved.

The goal is to show that \( g(x) = f(x) - l(x) < 0 \) on \((a, b)\), since this will show that \( f(x) < l(x) \) on \((a, b)\). Since \( f \) and \( l \) are both continuous on \([a, b]\) then so is \( g \). Hence \( g \) has a global maximum somewhere in \([a, b]\), by the Extreme Value Theorem. Suppose the global maximum occurs at an interior point \( x = c \), i.e. for some \( c \) in the open interval \((a, b)\). Then \( g'(c) = 0 \) and \( g''(c) = f''(c) - l''(c) = f''(c) > 0 \), since \( l(x) \) is a line and hence has a second derivative of 0 for all \( x \). Then by the Second Derivative Test \( g \) has a local minimum at \( x = c \), which contradicts \( g \) having a global maximum at \( x = c \). Thus, the global maximum of \( g \) cannot occur at an interior point, so it must occur at one of the end points \( x = a \) or \( x = b \). In other words, either \( g(x) < g(a) \) or \( g(x) < g(b) \) for all \( x \) in \((a, b)\). But \( f(a) = l(a) \) and \( f(b) = l(b) \), so \( g(a) = 0 = g(b) \). Hence, \( g(x) < 0 \) for all \( x \) in \((a, b)\), i.e. \( f(x) < l(x) \) for all \( x \) in \((a, b)\).

Points where the concavity of a function changes have a special name:

A function \( f \) has an **inflection point** at \( x = c \) if the concavity of \( f \) changes around \( x = c \). That is, the function goes from concave up to concave down, or vice versa.

Note that to be an inflection point it does not suffice for the second derivative to be 0 at that point; the second derivative must change sign around that point, either from positive to negative or from negative to positive. For example, \( f(x) = x^3 \) has an inflection point at \( x = 0 \), since \( f''(x) = 6x < 0 \) for \( x < 0 \) and \( f''(x) = 6x > 0 \) for \( x > 0 \), i.e. \( f''(x) \) changes sign around \( x = 0 \) (and of course \( f''(0) = 0 \)). But for \( f(x) = x^4 \), \( x = 0 \) is **not** an inflection point even though \( f''(0) = 0 \), since \( f''(x) = 12x^2 \geq 0 \) is always nonnegative. That is, \( f(x) = x^4 \) is always concave up. Figure 4.2.3 below shows the difference:
Figure 4.2.3 Inflection vs non-inflection point at $x = 0$ with $f''(0) = 0$

Figure 4.2.3(b) shows that a point where the second derivative is 0 is a possible inflection point, but you still must check that the second derivative changes sign around that point. Using local minima and maxima, concavity and inflection points, and where a function increases or decreases, you can sketch the graph of a function.

**Example 4.8**

Sketch the graph of $f(x) = x^3 - 6x^2 + 9x + 1$. Find all local maxima and minima, inflection points, where the function is increasing or decreasing, and where the function is concave up or concave down.

**Solution:** Since $f'(x) = 3x^2 - 12x + 9 = 3(x - 1)(x - 3)$ then $x = 1$ and $x = 3$ are the only critical points. And since $f''(x) = 6x - 12$ then $f''(1) = -6 < 0$ and $f''(3) = 6 > 0$. So by the Second Derivative Test, $f$ has a local maximum at $x = 1$ and a local minimum at $x = 3$. Since $f''(x) = 6x - 12 < 0$ for $x < 2$ and $f''(x) = 6x - 12 > 0$ for $x > 2$, then $x = 2$ is an inflection point, and $f$ is concave down for $x < 2$ and concave up for $x > 2$. The table below shows where $f$ is increasing and decreasing, based on the sign of $f'$:

<table>
<thead>
<tr>
<th>$x$ values</th>
<th>$(x - 1)$</th>
<th>$(x - 3)$</th>
<th>$f'(x)$</th>
<th>direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x &lt; 1$</td>
<td>$-$</td>
<td>$-$</td>
<td>$+$</td>
<td>$f$ is increasing</td>
</tr>
<tr>
<td>$1 &lt; x &lt; 3$</td>
<td>$+$</td>
<td>$-$</td>
<td>$-$</td>
<td>$f$ is decreasing</td>
</tr>
<tr>
<td>$x &gt; 3$</td>
<td>$+$</td>
<td>$+$</td>
<td>$+$</td>
<td>$f$ is increasing</td>
</tr>
</tbody>
</table>

The graph is shown below:
Example 4.9

Sketch the graph of \( f(x) = \frac{-x}{1+x^2} \). Find all local maxima and minima, inflection points, where the function is increasing or decreasing, and where the function is concave up or concave down. Also indicate any asymptotes.

Solution: Since \( f'(x) = \frac{x^2 - 1}{(1+x^2)^2} \) then \( x = 1 \) and \( x = -1 \) are the only critical points. And since \( f''(x) = \frac{2x(3-x^2)}{(1+x^2)^3} \) then \( f''(1) = \frac{1}{2} > 0 \) and \( f''(-1) = -\frac{1}{2} < 0 \). So by the Second Derivative Test, \( f \) has a local minimum at \( x = 1 \) and a local maximum at \( x = -1 \). Since \( f''(x) > 0 \) for \( x < -\sqrt{3} \), \( f''(x) < 0 \) for \( -\sqrt{3} < x < 0 \), \( f''(x) > 0 \) for \( 0 < x < \sqrt{3} \), and \( f''(x) < 0 \) for \( x > \sqrt{3} \), then \( x = 0, \pm \sqrt{3} \) are inflection points, \( f \) is concave up for \( x < -\sqrt{3} \) and for \( 0 < x < \sqrt{3} \), and \( f \) is concave down for \( -\sqrt{3} < x < 0 \) and for \( x > \sqrt{3} \). Since \( f'(x) > 0 \) for \( x < -1 \) and \( x > 1 \) then \( f \) is increasing for \( |x| > 1 \). And \( f'(x) < 0 \) for \( -1 < x < 1 \) means \( f \) is decreasing for \( |x| < 1 \). Finally, since \( \lim_{x \to \infty} f(x) = 0 \) and \( \lim_{x \to -\infty} f(x) = 0 \) then the \( x \)-axis (\( y = 0 \)) is a horizontal asymptote. There are no vertical asymptotes (why?). The graph is shown below:

![Graph of f(x) = -x/(1+x^2)](image)

If the Second Derivative Test fails then one alternative is the following test:

**First Derivative Test:** For a continuous function \( f \) on an interval \( I \), let \( x = c \) be a number in \( I \) such that \( f(c) \) is defined, and either \( f'(c) = 0 \) or \( f'(c) \) does not exist. Then:

(a) If \( f'(x) \) changes from negative to positive around \( x = c \) then \( f \) has a local minimum at \( x = c \).

(b) If \( f'(x) \) changes from positive to negative around \( x = c \) then \( f \) has a local maximum at \( x = c \).

This test merely states the obvious: a function decreases then increases around a minimum, and it increases then decreases around a maximum.
Example 4.10

Sketch the graph of \( f(x) = x^{2/3} \).

Solution: Clearly \( f(x) \) is continuous for all \( x \), including \( x = 0 \) (since \( f(0) = 0 \)), but \( f'(x) = \frac{2}{3}x^{-1/3} \) is not defined at \( x = 0 \). Since \( f'(x) \) changes from negative to positive around \( x = 0 \) \((f'(x) < 0 \text{ when } x < 0 \text{ and } f'(x) > 0 \text{ when } x > 0)\), then by the First Derivative Test \( f \) has a local minimum at \( x = 0 \). Since \( f''(x) = -\frac{2}{9}x^{-2/3} < 0 \) for all \( x \neq 0 \), then \( f \) is always concave down. There are no vertical or horizontal asymptotes. The graph is shown on the right.

Note that the Second Derivative Test could not be used for this function, since \( f'(x) \neq 0 \) for all \( x \) (notice also that \( f''(x) \) is not defined at \( x = 0 \)).

A more complete alternative to the Second Derivative Test is the following:\(^4\)

\textbf{\( N^{\text{th}} \) Derivative Test:} A function \( f \) with continuous derivatives of all orders up to and including \( n > 1 \) at \( x = c \) has either a local minimum, local maximum or inflection point at \( x = c \) if and only if

\[ f^{(k)}(c) = 0 \text{ for } k = 1, 2, \ldots, n-1 \text{ and } f^{(n)}(c) \neq 0 \]

(i.e. the \( n^{\text{th}} \) derivative is the first nonzero derivative at \( x = c \)). If so, then:

(a) If \( n > 1 \) is even and \( f^{(n)}(c) > 0 \) then \( f \) has a local minimum at \( x = c \).

(b) If \( n > 1 \) is even and \( f^{(n)}(c) < 0 \) then \( f \) has a local maximum at \( x = c \).

(c) If \( n > 1 \) is odd then \( f \) has an inflection point at \( x = c \).

Note that the Second Derivative Test is the special case where \( n = 2 \) in the \( N^{\text{th}} \) Derivative Test. Though this test gives necessary and sufficient conditions for a local maximum, local minimum, and inflection point, calculating the first \( n \) derivatives can be complicated if \( n \) is large and the given function is not simple.

Example 4.11

The Second Derivative Test fails for \( f(x) = x^4 \) at the critical point \( x = 0 \), since \( f''(0) = 0 \). But the first 4 derivatives of \( f(x) = x^4 \) are \( f'(x) = 4x^3 \), \( f''(x) = 12x^2 \), \( f^{(3)}(x) = 24x \), and \( f^{(4)}(x) = 24 \), which are all continuous and

\[ f^{(k)}(0) = 0 \text{ for } k = 1, 2, 3 \text{ and } f^{(4)}(0) = 24 \neq 0 . \]

So by the \( N^{\text{th}} \) Derivative Test, since \( n = 4 \) is even and \( f^{(4)}(0) = 24 > 0 \) then \( f(x) = x^4 \) has a local minimum at \( x = 0 \). Note that \( f(x) \geq 0 = f(0) \) for all \( x \), so \( x = 0 \) is actually a global minimum for \( f \).

A common practice in many fields of science and engineering is to combine multiple named constants (e.g. $\pi$) or variables in a function into one variable and then sketch a graph of that function. The example below illustrates the technique.

**Example 4.12**

A hydrogen atom has one electron, and the probability of finding the electron in the ground state of the hydrogen atom between radii $r$ and $r + dr$ is $D(r)dr$, where $dr$ is an infinitesimal change in the radius $r$ (the distance from the electron to the nucleus), $D(r)$ is the radial probability density function

$$D(r) = \frac{4}{a_0^3} r^2 e^{-2r/a_0}$$

and $a_0 \approx 5.291772 \times 10^{-11}$ m is the Bohr radius. It is useful to analyze this function in terms of $r \geq 0$ in relation to the Bohr radius $r_0$ (e.g. $r = 0.5a_0, a_0, 2a_0, 3a_0$). To do this, let $x = \frac{r}{a_0}$, so that

$$D(r) = \frac{4}{a_0^3} \left( \frac{r}{a_0} \right)^2 e^{-2\left( \frac{r}{a_0} \right)}$$

$$\Rightarrow a_0 D(x) = 4x^2 e^{-2x}$$

and then sketch the graph of $a_0 D(x)$, which is shown below:

![Graph of $a_0 D(x)$](image)

From the graph it looks like $x = 1$ (i.e. $r = a_0$) is a local (and global) maximum, so that the electron is most likely to be found near $r = a_0$, and the probability drops off dramatically past a distance $r = 3a_0$.

In the exercises you will be asked to show that $r = a_0$ is indeed a local maximum and that the inflection points are $r = \left( 1 \pm \frac{1}{\sqrt{2}} \right) a_0$.

Note that the right side of the formula $a_0 D(x) = 4x^2 e^{-2x}$ does not involve $a_0$, which was multiplied over to the left side. In general that is the strategy when dealing with these sorts of functions where variables and constants are combined. In this case the stray constant $a_0$ can be multiplied with $D$ since that will not affect the location of critical and inflection points, nor fundamentally alter the general shape of the graph.
Example 4.13

For a single particle with two states—energy 0 and energy $\epsilon$—in thermal contact with a reservoir at temperature $\tau$, the average energy $U$ and heat capacity $C_V$ are given by

$$U = \epsilon \frac{e^{-\epsilon/\tau}}{1 + e^{-\epsilon/\tau}}$$
$$C_V = k_B \left( \frac{\epsilon}{\tau} \right)^2 \frac{e^{\epsilon/\tau}}{(1 + e^{\epsilon/\tau})^2}$$

where $k_B \approx 1.38065 \times 10^{-23}$ J/K is the Boltzmann constant. The graph below shows both quantities as functions of $\tau/\epsilon$ (not $\epsilon/\tau$, as you might expect). See Exercise 9.

Average Energy $U/\epsilon$ vs Heat Capacity $C_V/k_B$

Exercises

A

For Exercises 1-8 sketch the graph of the given function. Find all local maxima and minima, inflection points, where the function is increasing or decreasing, where the function is concave up or concave down, and indicate any asymptotes.

1. $f(x) = x^3 - 3x$
2. $f(x) = x^3 - 3x^2 + 1$
3. $f(x) = xe^{-x}$
4. $f(x) = x^2 e^{-x^2}$
5. $f(x) = \frac{1}{1 + x^2}$
6. $f(x) = \frac{x^2}{(x - 1)^2}$
7. $f(x) = \frac{e^{-x} - e^{-2x}}{2}$
8. $f(x) = e^{-x} \sin x$

9. Write $U/\epsilon$ and $C_V/k_B$ from Example 4.13 as functions of $x = \tau/\epsilon$. You do not need to sketch the graphs.
10. Show that the function $D(r) = \frac{4}{a_0^2} r^2 e^{-2r/a_0}$ from Example 4.12 has a local maximum at $r = a_0$ and inflection points at $r = \left(1 \pm \frac{1}{\sqrt{2}}\right) a_0$.
11. Sketch the graph of Kratzer’s molecular potential $V(r) = -2D \left( \frac{a}{r} - \frac{1}{2} \frac{a^2}{r^2} \right)$ as a function of $x = \frac{r}{a}$, with $a > 0$ and $D > 0$ as constants.
12. Sketch the graph of $f(K) = \frac{2N\sqrt{\frac{K}{\tau}} e^{K/\sqrt{\tau(kT)^{3/2}}}}{\sqrt{\tau(kT)^{3/2}}}$ as a function of $x = \frac{K}{kT}$, with $N$, $k$ and $T$ as positive constants.
13. Prove part(b) of the Concavity Theorem.
4.3 Numerical Approximation of Roots of Functions

When finding critical points of a function $f$, you encounter the problem of solving the equation $f'(x) = 0$. The examples and exercises so far were set up carefully so that solutions to that equation could be found in a simple closed form. But in practice this will not always be the case—in fact it is almost never the case. For example, finding the critical points of $f(x) = \sin x - \frac{x^2}{2}$ entails solving the equation $f'(x) = \cos x - x = 0$, for which there is no solution in a closed-form expression.

What should you do in such a situation? One possibility is to use the bisection method mentioned in Section 3.3. In fact, in Example 3.25 the solution to the equation $\cos x = x$ (i.e. $\cos x - x = 0$) was shown to exist in the interval $[0, 1]$, and then a demonstration of the bisection method was given to find that solution.

The bisection method is one of many numerical methods for finding roots of a function (i.e. where the function is zero). Finding the critical points of a function means finding the roots of its derivative. Though the bisection method could be used for that purpose, it is not efficient—convergence to the root is slow. A far more efficient method is Newton’s method, whose geometric interpretation is shown in Figure 4.3.1 below.

![Figure 4.3.1](image)

Newton’s method for finding a root $\bar{x}$ of $f(x)$

The idea behind Newton’s method is simple: to find a root $\bar{x}$ of a function $f$, choose an initial guess $x_0$ and then go up—or down—to the curve $y = f(x)$ and draw the tangent line to the curve at the point $(x_0, f(x_0))$. Let $x_1$ be where that tangent line intersects the $x$-axis, as shown above; repeat this procedure on $x_1$ to get the next number $x_2$, repeat on $x_2$ to get $x_3$, and so on. The resulting sequence of numbers $x_0, x_1, x_2, x_3, \ldots$, will approach the root $\bar{x}$. Convergence under certain conditions can be proved.²

²Note: To “just give up”—as suggested semi-seriously by some students I have had—is not an option.

²Sometimes called the Newton-Raphson method.

The general formula for the number $x_n$ obtained after $n \geq 1$ iterations in Newton’s method can be determined by considering the formula for $x_1$. First, the tangent line to $y = f(x)$ at the point $(x_0, f(x_0))$ has slope $f'(x_0)$, so the equation of the line is

$$y - f(x_0) = f'(x_0)(x - x_0).$$

The point $(x_1, 0)$ is (by design) also on that line, so that

$$0 - f(x_0) = f'(x_0)(x_1 - x_0) \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

provided that $f'(x_0) \neq 0$. The general formula for $x_n$ is given by the following algorithm:

**Newton’s method**: For an initial guess $x_0$, the numbers $x_n$ for $n \geq 1$ are computed iteratively as:

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \quad \text{for } n = 1, 2, 3, \ldots$$

That is, each “next” number $x_n$ depends on the previous number $x_{n-1}$. The algorithm terminates whenever $f'(x_n) = 0$, or when the desired accuracy is reached. If $f'(x_n) = 0$ for some $n \geq 0$, then you could start over with a different initial guess $x_0$.

To implement this algorithm in a programming language (for which Newton’s method is well-suited), the following language-independent pseudocode can be used as a guide:

```plaintext
Algorithm pseudocode for Newton's method

Newton's method

N ← NUMBER-OF-ITERATIONS       ▶ User supplies this value
x ← INITIAL-GUESS              ▶ User supplies this value
for n ← 1 to N
  do
    if f'(x) ≠ 0
      then
        x ← x - \frac{f(x)}{f'(x)}
        print x
      else
        error "division by zero"
  done
```

```
Example 4.14

Use Newton's method to find the root of \( f(x) = \cos x - x \).

**Solution:** Since the root is already known to be in the interval \([0,1]\), choose \(x_0 = 1\) as the initial guess. The numbers \(x_n\) for \(n \geq 1\) can be computed with a hand-held scientific calculator, but the process is tedious and error-prone. Using a computer is far more efficient and allows more flexibility.

For example, the algorithm is easily implemented in the Java programming language. Save this code in a plain text file as `newton.java`:

```
1 public class newton {
2   public static void main(String[] args) {
3     int N = Integer.parseInt(args[0]); // Number of iterations
4     double x = 1.0; // initial guess
5     System.out.println("n=0: " + x);
6     for (int i = 1; i <= N; i++) {
7       x = x - f(x)/derivf(x);
8       System.out.println("n=" + i + ": " + x);
9     }
10   }
11 }
12
13 // Define the function f(x)
14 public static double f(double x) {
15   return Math.cos(x) - x;
16 }
17
18 // Define the derivative f'(x)
19 public static double derivf(double x) {
20   return -Math.sin(x) - 1.0;
21 }
22
23 Though knowledge of Java would help, it should not be that difficult to figure out what the above code is doing. The number of iterations \(N\) is passed as a command-line parameter to the program, and \(x_n\) is computed and printed for \(n = 0, 1, 2, \ldots, N\). Note that the derivative of \(f(x)\) is “hard-coded” into the program.\(^8\) There is also no error checking for the derivative being zero at any \(x_n\). The program would simply halt on a division by zero error.

Compile the code, then run the program with 10 iterations:

```
javac newton.java
java newton 10
```

The output is shown below:

```
8There are some programming language libraries for calculating derivatives of functions “on the fly,” i.e. dynamically. For example, the GNU libmatheval C/Fortran library can perform such symbolic operations. It is available at http://www.gnu.org/software/libmatheval/
```
Numerical Approximation of Roots of Functions • Section 4.3

n=0: 1.0  
n=1: 0.7503638678402439  
n=2: 0.7391128909113617  
n=3: 0.739085133385284  
n=4: 0.7390851332151607  
n=5: 0.7390851332151607  
n=6: 0.7390851332151607  
n=7: 0.7390851332151607  
n=8: 0.7390851332151607  
n=9: 0.7390851332151607  
n=10: 0.7390851332151607

Note that the solution $\bar{x} = 0.7390851332151607$ was found after only 4 iterations; the numbers $x_n$ repeat for $n \geq 5$. This is much faster than the bisection method.

Another root-finding numerical method similar to Newton’s method is the **secant method**, whose geometric interpretation is shown in Figure 4.3.2 below:

![Figure 4.3.2 Secant method for finding a root $\bar{x}$ of $f(x)$](image)

The idea behind the secant method is simple: to find a root $\bar{x}$ of a function $f$, choose *two* initial guesses $x_0$ and $x_1$, then go up—or down—to the curve $y = f(x)$ and draw the secant line through the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$ on the curve. Let $x_2$ be where that secant line intersects the $x$-axis, as shown above; repeat this procedure on $x_1$ and $x_2$ to get the next number $x_3$, and keep repeating in this way. The resulting sequence of numbers $x_0, x_1, x_2, x_3, \ldots$, will approach the root $\bar{x}$, under the right conditions.\(^9\)

Since the secant line through \((x_0, f(x_0))\) and \((x_1, f(x_1))\) has slope \(\frac{f(x_1) - f(x_0)}{x_1 - x_0}\), the equation of that secant line is:

\[
y - f(x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x - x_1)
\]

The point \((x_2, 0)\) is on that line, so that

\[
0 - f(x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x_2 - x_1) \Rightarrow x_2 = x_1 - \frac{(x_1 - x_0) \cdot f(x_1)}{f(x_1) - f(x_0)}
\]

provided that \(x_1 \neq x_0\). The general formula for \(x_n\) is given by the following algorithm:

**Secant method**: For two initial guesses \(x_0\) and \(x_1\), the numbers \(x_n\) for \(n \geq 2\) are computed iteratively as:

\[
x_n = x_{n-1} - \frac{(x_{n-1} - x_{n-2}) \cdot f(x_{n-1})}{f(x_{n-1}) - f(x_{n-2})} \quad \text{for } n = 2, 3, 4, \ldots
\]

That is, each “next” number \(x_n\) depends on the previous two numbers \(x_{n-1}\) and \(x_{n-2}\). The algorithm terminates whenever \(x_n = x_{n-1}\) (i.e. the numbers start repeating) or when the desired accuracy is reached.

**Algorithm pseudocode for the secant method**

```
SECANT METHOD
1  N ← NUMBER-OF-ITERATIONS  ▷ User supplies this value
2  x0 ← FIRST-INITIAL-GUESS   ▷ User supplies this value
3  x1 ← SECOND-INITIAL-GUESS ▷ User supplies this value
4  f0 ← f(x0)
5  for n ← 1 to N
6    do
7      fn ← f(xn)
8      if f0 ≠ fn
9        then
10         x ← x1 - \frac{(x_1 - x_0) \cdot f_1}{f_1 - f_0}
11         print x
12         x0 ← x1
13         f0 ← fn  ▷ Re-use f_1 as f_0 in the next iteration
14         x1 ← x
15      else error “division by zero”
```
One difference you might have noticed between the secant method and Newton’s method is that the secant method does not use derivatives. The secant method replaces the derivative in Newton’s method with the slope of a secant line which approximates the derivative (recall how the tangent line is the limit of slopes of secant lines). This might seem like a drawback, perhaps giving a “less accurate” slope than the tangent line, but in practice it is not really a problem. In fact, in many cases the secant method is preferable, since computing derivatives can often be quite complicated.

Example 4.15

Use the secant method to find the root of \( f(x) = \cos x - x \).

Solution: Since the root is already known to be in the interval \([0, 1]\), choose \( x_0 = 0 \) and \( x_1 = 1 \) as the two initial guesses. The algorithm is easily implemented in the Java programming language. Save this code in a plain text file as secant.java:

```
Listing 4.2 Secant method in Java (secant.java)

import java.math.*;
public class secant {
    public static void main(String[] args) {
        int N = Integer.parseInt(args[0]); // Number of iterations
        double x0 = 0.0; // first initial guess
        double x1 = 1.0; // second initial guess
        double f0 = f(x0);
        double f1;
        double x = 0.0;
        for (int i = 2; i <= N; i++) {
            f1 = f(x1);
            x = x1 - (x1 - x0)*f1/(f1 - f0);
            x0 = x1;
            f0 = f1; // Re-use f1 as f0 in the next iteration
            x1 = x;
            System.out.println("n=" + i + ": " + x);
        }
    }
}

// Define the function f(x)
public static double f(double x) {
    return Math.cos(x) - x;
}
```

Compile the code, then run the program:

```
javac secant.java
java secant 10
```

The output is shown below:
114 Chapter 4 • Applications of Derivatives §4.3

For the function $f(x) = \cos x - x$, the table below summarizes the results of 10 iterations of the bisection method, Newton’s method and the secant method:

<table>
<thead>
<tr>
<th>Term</th>
<th>Bisection</th>
<th>Newton</th>
<th>Secant</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_0$</td>
<td>0.5</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$x_1$</td>
<td>0.75</td>
<td>0.7503638678402439</td>
<td>1.0</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.625</td>
<td>0.7391128909113617</td>
<td>0.685073573260451</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.6875</td>
<td>0.739085133385284</td>
<td>0.736298997613654</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.71875</td>
<td>0.7390851332151607</td>
<td>0.7391193619116293</td>
</tr>
<tr>
<td>$x_5$</td>
<td>0.734375</td>
<td>0.7390851332151607</td>
<td>0.7390851121274639</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.7421875</td>
<td>0.7390851332151607</td>
<td>0.7390851332150012</td>
</tr>
<tr>
<td>$x_7$</td>
<td>0.73828125</td>
<td>0.7390851332151607</td>
<td>0.7390851332151607</td>
</tr>
<tr>
<td>$x_8$</td>
<td>0.740234375</td>
<td>0.7390851332151607</td>
<td>0.7390851332151607</td>
</tr>
<tr>
<td>$x_9$</td>
<td>0.7392578125</td>
<td>0.7390851332151607</td>
<td>undefined</td>
</tr>
<tr>
<td>$x_{10}$</td>
<td>0.73876953125</td>
<td>0.7390851332151607</td>
<td>undefined</td>
</tr>
</tbody>
</table>

Newton’s method found the root after 4 iterations, while the secant method needed 6 iterations. After 10 iterations the bisection method had yet to find the root to the same level of precision as the other methods—it would take 52 iterations (that is, $x_{52}$) to achieve similar accuracy to 16 decimal places.

In general Newton’s method requires fewer iterations to find a root than the secant method does, but this does not necessarily mean that it will always be faster. Depending on the complexity of the function and its derivative, Newton’s method could involve more “expensive” operations (i.e. computing values, as opposed to assigning...
values) than the secant method, so that the few more iterations possibly required by the secant method are made up for by fewer total computations.

To see this, notice that Newton’s method always requires that both \( f(x_{n-1}) \) and \( f'(x_{n-1}) \) be computed for the \( n^{th} \) term \( x_n \) in the sequence. The secant method needs \( f(x_{n-1}) \) and \( f(x_{n-2}) \) for the \( n^{th} \) term, but a good programmer would save the value of \( f(x_{n-1}) \) so that it could be re-used (and hence not re-computed) as \( f(x_{n-2}) \) in the next iteration, resulting in potentially fewer total computations for the secant method.

There are occasional pitfalls in using Newton’s method. For example, if you were unlucky enough to reach a number \( x_n \) such that \( f'(x_n) = 0 \), then Newton’s method fails, due to division by 0 in the algorithm. Geometrically the reason is clear: the tangent line to the curve at that point would be parallel to the \( x \)-axis and hence would not intersect it (assuming \( f(x_n) \neq 0 \)). There would be no “next number” \( x_{n+1} \) in the iteration! See Figure 4.3.3(a) below.

Another possible problem is that Newton’s method might move you away from the root, i.e. not get closer, typically by a poor choice of \( x_0 \). See Figure 4.3.3(b) above. In some extreme cases, it is possible that Newton’s method simply loops back and forth endlessly between the same two numbers, as in Figure 4.3.4:

In most cases a different choice for the initial guess \( x_0 \) will fix such problems.
Most textbooks on the subject of numerical analysis discuss these issues.\textsuperscript{10} There are conditions under which Newton’s method is guaranteed to work, and convergence is fast. Newton’s method has a \textit{quadratic} rate of convergence, meaning roughly that the error terms—the differences between approximate roots and the actual root—are being squared in the long term. More precisely, if the numbers $x_n$ for $n \geq 0$ converge to a root $\bar{x}$, then the error terms $\epsilon_n = x_n - \bar{x}$ for $n \geq 0$ satisfy the limit
\[
\lim_{n \to \infty} \frac{|\epsilon_{n+1}|}{|\epsilon_n|^2} = C
\]
for some constant $C$. Squared error terms might sound like a bad thing, but the $x_n$ terms are converging to the root, making the error terms closer to 0 for large $n$. Squaring a number $\epsilon_n$ when $|\epsilon_n| < 1$ results in a much smaller number, not a larger one.

Using a computer to see the graph of the function often makes the choice of the initial guess $x_0$ easier and less prone to problems. By the way, these numerical methods allow you to sketch the graphs of many more functions, since finding local minima and maxima involves finding roots of $f’$, and finding inflection points involves finding roots of $f’’$. You now know some methods for finding those roots.

Finally, despite being much slower, the bisection method has the enormous advantage of always working. With the speed of modern computers the difference in algorithmic efficiency could be negligible in many cases.

\section*{Exercises}

\begin{enumerate}
\item Use Newton’s method to find the root of $f(x) = \cos x - 2x$.
\item Use Newton’s method to find the positive root of $f(x) = \sin x - x/2$.
\item Use Newton’s method to find the solution of the equation $e^{-x} = x$.
\item Use Newton’s method to find the solution of the equation $e^{-x} = x^2$.
\item Use Newton’s method and $f(x) = x^2 - 2$ to approximate $\sqrt{2}$ accurate to six decimal places.
\item Use Newton’s method to approximate $\sqrt{3}$ accurate to six decimal places.
\item Repeat Exercise 1 with the secant method.
\item Repeat Exercise 3 with the secant method.
\item Repeat Exercise 5 with the secant method.
\item Repeat Exercise 6 with the secant method.
\item Cosmic microwave background radiation is described by a function similar to $f(x) = \frac{x^3}{1 + x^2}$ for $x \geq 0$. Use Newton’s method to find the global maximum of $f$ accurate to four decimal places.
\item Would a different choice for $x_0$ in either graph in Figure 4.3.4 eliminate the infinite loop? Explain.
\item Draw a graph without any symmetry that has the infinite loop problem for Newton’s method.
\end{enumerate}

4.4 The Mean Value Theorem

The difference between instantaneous and average rates of change has been discussed in earlier sections. Recall that there is no difference between the two for linear functions. For nonlinear functions the average rate of change over an interval \([a, b]\) of positive length (i.e. \(b - a > 0\)) will not be the same as the instantaneous rate of change at every point in the interval. However, the following theorem guarantees that they will be the same at some point in the interval:

**Mean Value Theorem:** Let \(a\) and \(b\) be real numbers such that \(a < b\), and suppose that \(f\) is a function such that

(a) \(f\) is continuous on \([a, b]\), and

(b) \(f\) is differentiable on \((a, b)\).

Then there is at least one number \(c\) in the interval \((a, b)\) such that

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

Figure 4.4.1 below shows the geometric interpretation of the theorem:

![Figure 4.4.1](image)

**Figure 4.4.1** Mean Value Theorem: parallel tangent line and secant line

The idea is that there is at least one point on the curve \(y = f(x)\) where the tangent line will be parallel to the secant line joining the points \((a, f(a))\) and \((b, f(b))\). For each \(c\) in \((a, b)\) the tangent line has slope \(f'(c)\), while the secant line has slope \(\frac{f(b) - f(a)}{b - a}\). The Mean Value Theorem says that these two slopes will be equal somewhere in \((a, b)\).

To prove the Mean Value Theorem (sometimes called *Lagrange’s Theorem*), the following intermediate result is needed, and is important in its own right:
Rolle’s Theorem: Let \(a\) and \(b\) be real numbers such that \(a < b\), and suppose that \(f\) is a function such that

(a) \(f\) is continuous on \([a, b]\),

(b) \(f\) is differentiable on \((a, b)\), and

(c) \(f(a) = f(b) = 0\).

Then there is at least one number \(c\) in the interval \((a, b)\) such that \(f'(c) = 0\).

Figure 4.4.2 on the right shows the geometric interpretation of the theorem. To prove the theorem, assume that \(f\) is not the constant function \(f(x) = 0\) for all \(x\) in \([a, b]\) (if it were then Rolle’s Theorem would hold trivially). Then there must be at least one \(x_0\) in \((a, b)\) such that either \(f(x_0) > 0\) or \(f(x_0) < 0\). If \(f(x_0) > 0\) then by the Extreme Value Theorem \(f\) attains a global maximum at some \(x = c\) in the open interval \((a, b)\), since \(f\) is zero at the endpoints \(x = a\) and \(x = b\) of the closed interval \([a, b]\). Then \(f'(c) = 0\) since \(f\) has a maximum at \(x = c\). Likewise if \(f(x_0) < 0\) then \(f\) attains a global minimum at some \(x = c\) in \((a, b)\), and thus again \(f'(c) = 0\).

The Mean Value Theorem can now be proved by applying Rolle’s Theorem to the function

\[
F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)
\]

where \(f\) satisfies the conditions of the Mean Value Theorem. Basically, \(F\) “tilts” the graph of \(f\) from Figure 4.4.1 to look like the graph in Figure 4.4.2. It is trivial to check that \(F(a) = F(b) = 0\), and \(F\) is continuous on \([a, b]\) and differentiable on \((a, b)\) since \(f\) is. Thus, by Rolle’s Theorem, \(F'(c) = 0\) for some \(c\) in \((a, b)\). However,

\[
F'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}
\]

and so

\[
F'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0 \implies f'(c) = \frac{f(b) - f(a)}{b - a}
\]

Note that both the Mean Value Theorem and Rolle’s Theorem are purely existence theorems—they tell you only that a certain number exists. The task of finding the numbers is left to you. For the Mean Value Theorem that task involves solving the equation \(f'(x) = \frac{f(b) - f(a)}{b - a}\) (or \(f'(x) = 0\) for Rolle’s Theorem). The numerical root-finding methods from Section 4.3 could come in handy, since obtaining closed-form solutions might be impossible. For that reason, the Mean Value Theorem is more useful for theoretical purposes. One such application is the following important result:
If $f$ is a differentiable function on an interval $I$ such that $f'(x) = 0$ for all $x$ in $I$, then $f$ is a constant function on $I$.

Note that $I$ can be any interval, even the entire real line $(-\infty, \infty)$. It is already known that $f = \text{constant} \Rightarrow f' = 0$; the above result says that the converse is true. The proof is by contradiction: assume that $f$ is not a constant function and show this contradicts the Mean Value Theorem. If $f$ is not constant then there exist numbers $a < b$ in $I$ such that $f(a) \neq f(b)$. However, by the Mean Value Theorem there must exist a number $c$ in the interval $(a, b)$ such that

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

Since the derivative of $f$ is 0 everywhere in $I$, then $f'(c) = 0$ and so

$$\frac{f(b) - f(a)}{b - a} = 0 \Rightarrow f(b) - f(a) = 0 \Rightarrow f(a) = f(b),$$

a contradiction of $f(a) \neq f(b)$. Thus, $f$ must be a constant function. ✓

Another theoretical result can be proved with the Mean Value Theorem:

Let $f$ be a differentiable function on an interval $I$. Then:

(a) If $f' > 0$ on $I$ then $f$ is increasing on $I$.

(b) If $f' < 0$ on $I$ then $f$ is decreasing on $I$.

To prove part(a), assume that $f'(x) > 0$ for all $x$ in $I$, and choose arbitrary numbers $a$ and $b$ in $I$ with $a < b$. To prove that $f$ is increasing on $I$ it suffices to show that $f(a) < f(b)$. By the Mean Value Theorem there is a number $c$ in $(a, b)$ (and hence in $I$) such that

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

and so

$$f(b) - f(a) = (b - a)f'(c) > 0$$

since $b - a > 0$ and $f'(c) > 0$. Thus, $f(b) > f(a)$, and so $f$ is increasing on $I$. ✓

The proof of part(b) is similar and is left as an exercise. You might wonder why such a proof is necessary. After all, an intuitive explanation was provided in Section 1.2 for why positive or negative derivatives imply that a function is increasing or decreasing, respectively. That knowledge has been assumed and used in the subsequent sections. Intuitive so-called “hand-waving” explanations, in fact, often yield more insight than a “formal” proof, such as the one above. However, it is good to know that such intuition has a solid basis and can be proved, if needed.
The Mean Value Theorem can help in proving inequalities, often used in the sciences for establishing upper or lower bounds on a quantity (e.g. worst-case scenario).

**Example 4.16**

Show that \( \sin x \leq x \) for all \( x \geq 0 \).

**Solution:** The inequality holds trivially for \( x = 0 \), since \( \sin 0 = 0 \). So assume that \( x > 0 \). Then by the Mean Value Theorem there is a number \( c \) in \((0, x)\) such that for \( f(x) = \sin x \),

\[
\frac{f(x) - f(0)}{x - 0} = f'(c) \implies \frac{\sin x - \sin 0}{x - 0} = \cos c
\]

\[
\Rightarrow \sin x = x \cos c
\]

\[
\Rightarrow \sin x \leq x
\]

since \( \cos c \leq 1 \) and \( x > 0 \). Note that \( \sin x \leq x \) is a sharper inequality than \( \sin x \leq 1 \) when \( 0 < x < 1 \).

There is a useful alternative form of the Mean Value Theorem. If \( a < b \) then let \( h = b - a > 0 \), so that \( a + h = b \). Then any number \( c \) in \((a, b)\) can be written as \( c = a + \theta h \) for some number \( \theta \) in \((0, 1)\). To see this, let \( c \) be in \((a, b)\). Then \( 0 < c - a < b - a = h \) and so \( 0 < \frac{c - a}{h} < 1 \). Thus, \( \theta = \frac{c - a}{h} \) is in \((0, 1)\) and \( a + \theta h = a + (c - a) = c \). Hence:

**Mean Value Theorem (alternative form):** Let \( a \) and \( h > 0 \) be real numbers, and suppose that \( f \) is a function such that

(a) \( f \) is continuous on \([a, a + h]\), and

(b) \( f \) is differentiable on \((a, a + h)\).

Then there is a number \( \theta \) in the interval \((0, 1)\) such that

\[
f(a + h) - f(a) = h f'(a + \theta h).
\]

The Mean Value Theorem is the special case of \( g(x) = x \) in the following generalization:

**Extended Mean Value Theorem:** Let \( a \) and \( b \) be real numbers such that \( a < b \), and suppose that \( f \) and \( g \) are functions such that

(a) \( f \) and \( g \) are continuous on \([a, b]\),

(b) \( f \) and \( g \) are differentiable on \((a, b)\), and

(c) \( g'(x) \neq 0 \) for all \( x \) in \((a, b)\).

Then there is at least one number \( c \) in the interval \((a, b)\) such that

\[
\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.
\]
The Mean Value Theorem says that the derivative of a differentiable function will always attain one particular value on a closed interval: the function’s average rate of change over the interval. It turns out that the derivative will take on every value between its values at the endpoints, similar to how the Intermediate Value Theorem applies to continuous functions:

**Darboux’s Theorem:** If \( f \) is a differentiable function on a closed interval \([a, b]\) then its derivative \( f'\) attains every value between \( f'(a) \) and \( f'(b) \).

In other words, if \( f'(a) < \gamma < f'(b) \) (or \( f'(b) < \gamma < f'(a) \)) then there is a number \( c \) in \((a, b)\) such that \( f'(c) = \gamma \). If \( f' \) were continuous on \([a, b]\) then the result would follow trivially by the Intermediate Value Theorem for continuous functions. What is perhaps surprising is that Darboux’s Theorem holds even for derivatives that are not continuous. This means that a discontinuous derivative cannot have the type of simple jump discontinuities that would allow it to “skip” over intermediate values—the points of discontinuity must be of a more complicated type. One rough interpretation of Darboux’s Theorem is that even if a derivative is not a continuous function, it will behave sort of as if it were.

### Exercises

**A**

1. Does Rolle’s Theorem apply to the function \( f(x) = 1 - |x| \) on the interval \([-1, 1]\)? If so, find the number in \((-1, 1)\) that Rolle’s Theorem guarantees to exist. If not, explain why not.

2. Suppose that two horses run a race starting together and ending in a tie. Show that, at some time during the race, they must have had the same speed.

3. Use the Mean Value Theorem to show that \(|\sin A - \sin B| \leq |A - B|\) for all \(A\) and \(B\) (in radians). Does \(|\sin A + \sin B| \leq |A + B|\) for all \(A\) and \(B\)? Explain.

4. Show that \(|\cos A - \cos B| \leq |A - B|\) for all \(A\) and \(B\) (in radians). Does \(|\cos A + \cos B| \leq |A + B|\) for all \(A\) and \(B\)? Explain.

5. Show that \(\tan x \geq x\) for all \(0 \leq x < \frac{\pi}{2}\).

6. Show that \(|\tan A - \tan B| \geq |A - B|\) for all \(A\) and \(B\) (in radians) in \((-\frac{\pi}{2}, \frac{\pi}{2})\). Can the inequality be extended to all \(A\) and \(B\)? Explain your answer.

**B**

7. Use Rolle’s Theorem to show that for all constants \(a\) and \(b\) with \(a > 0\), \(f(x) = x^3 - ax + b\) can not have three positive roots. Also, show that it can not have three negative roots.

---

8. Use the Mean Value Theorem to show that if \( f' < 0 \) on an interval \( I \) then \( f \) is decreasing on \( I \).

9. Suppose that \( f \) and \( g \) are continuous on \([a,b]\) and differentiable on \((a,b)\), and that \( f'(x) > g'(x) \) for all \( a < x < b \). Show that \( f(b) - g(b) > f(a) - g(a) \).

10. Prove the Extended Mean Value Theorem, by applying Rolle’s Theorem to the function

\[
F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a)).
\]

11. Show that \( e^x \geq 1 + x \) for all \( x \). (Hint: Consider \( f(x) = e^x - x \).)

12. Show that \( \ln(1 + x) < x \) for all \( x > 0 \).

13. Show that \( \tan^{-1} x < x \) for all \( x \).

14. Show that for \( 0 < a \leq \beta < \frac{\pi}{2} \),

\[
\frac{\beta - a}{\cos^2 a} \leq \tan \beta - \tan a \leq \frac{\beta - a}{\cos^2 \beta}.
\]

15. Show that for \( 0 < a \leq b \),

\[
\frac{b - a}{b} \leq \ln \frac{b}{a} \leq \frac{b - a}{a}.
\]

16. Show that for \( n > 1 \) and \( a > b \),

\[
nb^{n-1}(a - b) < a^n - b^n < na^{n-1}(a - b).
\]

C

17. Show that \( \sqrt{a^2 + b} < a + \frac{b}{2a} \) for all positive numbers \( a \) and \( b \).

18. Show that \( f(x) = \cos^2 x + \cos^2 \left( \frac{\pi}{4} + x \right) - \cos x \cos \left( \frac{\pi}{4} + x \right) \) is a constant function. What is its value?

19. Suppose that \( f(x) \) is a differentiable function and that \( f(0) = 0 \) and \( f(1) = 1 \). Show that \( f'(x_0) = 2x_0 \) for some \( x_0 \) in the interval \((0,1)\).

20. Prove the inequality

\[
\left| \frac{x_1 + x_2}{1 + x_1 x_2} \right| < 1 \quad \text{for} \quad -1 < x_1, x_2 < 1
\]

as follows:

(a) First prove the special case where \( x_1 = x_2 \).

(b) For the case \( x_1 < x_2 \) define

\[
f(x) = \frac{x + a}{1 + ax}
\]

for \(-1 \leq x \leq 1\), where \(-1 < a < 1\). Show that \( f \) is increasing on \([-1,1]\), then use \( a = x_2 \) and \( x = x_1 \).

Note that proving the case \( x_2 < x_1 \) is unnecessary (why?).

This inequality is a generalization of the same inequality for \( 0 \leq x_1, x_2 < 1 \) in the relativistic velocity addition law from the theory of special relativity: if object 1 has velocity \( v_1 \) relative to a frame of reference \( F \), and if object 2 has a velocity \( v_2 \) relative to object 1, so that \( x_1 = v_1/c \) and \( x_2 = v_2/c \) represent the fractions of the speed of light \( c \) at which the objects are moving, then the fraction of the speed of light at which object 2 is moving with respect to \( F \) is \( x = (x_1 + x_2)/(1 + x_1 x_2) \). So it should be true that \( 0 \leq x < 1 \), since nothing can move faster than the speed of light.
CHAPTER 5

The Integral

5.1 The Indefinite Integral

Derivatives appear in many physical phenomena, such as the motion of objects. Recall, for example, that given the position function $s(t)$ of an object moving along a straight line at time $t$, you could find the velocity $v(t) = s'(t)$ and the acceleration $a(t) = v'(t)$ of the object at time $t$ by taking derivatives. Suppose the situation were reversed: given the velocity function how would you find the position function, or given the acceleration function how would you find the velocity function?

![Figure 5.1.1](image)

Figure 5.1.1 Differentiation vs antidifferentiation for motion functions

In this case calculating a derivative would not help, since the reverse process is needed: instead of differentiation you need a way of performing antidifferentiation, i.e. you would calculate an antiderivative.

An antiderivative $F(x)$ of a function $f(x)$ is a function whose derivative is $f(x)$. In other words, $F'(x) = f(x)$.

Differentiation is relatively straightforward; you have learned the derivatives of many classes of functions (e.g. polynomials, trigonometric functions, exponential and logarithmic functions), and with the various rules for differentiation you can calculate derivatives of complicated expressions involving those functions (e.g. sums, powers, products, quotients). Antidifferentiation, however, is a different story.
To see some of the issues involved, consider a simple function like \( f(x) = 2x \). Of course you know that \( \frac{d}{dx}(x^2) = 2x \), so it seems that \( F(x) = x^2 \) is the antiderivative of \( f(x) = 2x \). But is it the only antiderivative of \( f(x) \)? No. For example, if \( F(x) = x^2 + 1 \) then \( F'(x) = 2x = f(x) \), and so \( F(x) = x^2 + 1 \) is another antiderivative of \( f(x) = 2x \). Likewise, so is \( F(x) = x^2 + 2 \). In fact, any function of the form \( F(x) = x^2 + C \), where \( C \) is some constant, is an antiderivative of \( f(x) = 2x \).

Another potential issue is that functions of the form \( F(x) = x^2 + C \) are just the most obvious antiderivatives of \( f(x) = 2x \). Could there be some other completely different function—one that cannot be simplified into the form \( x^2 + C \)—whose derivative also turns out to be \( f(x) = 2x \)? The answer, luckily, is no:

> Suppose that \( F(x) \) and \( G(x) \) are antiderivatives of a function \( f(x) \). Then \( F(x) \) and \( G(x) \) differ only by a constant. That is, \( F(x) = G(x) + C \) for some constant \( C \).

To prove this, consider the function \( H(x) = F(x) - G(x) \), defined for all \( x \) in the common domain \( I \) of \( F \) and \( G \). Since \( F'(x) = G'(x) = f(x) \), then

\[
H'(x) = F'(x) - G'(x) = f(x) - f(x) = 0
\]

for all \( x \) in \( I \), so \( H(x) \) is a constant function on \( I \), as was shown in Section 4.4 on the Mean Value Theorem. Thus, there is a constant \( C \) such that

\[
H(x) = C \quad \Rightarrow \quad F(x) - G(x) = C \quad \Rightarrow \quad F(x) = G(x) + C
\]

for all \( x \) in \( I \). ✓

The practical consequence of the above result can be stated as follows:

> To find all antiderivatives of a function, it is necessary only to find one antiderivative and then add a generic constant to it.

So for the function \( f(x) = 2x \), since \( F(x) = x^2 \) is one antiderivative then all antiderivatives of \( f(x) \) are of the form \( F(x) = x^2 + C \), where \( C \) is a generic constant. Thus, functions do not have just one antiderivative but a whole family of antiderivatives, all differing only by a constant. The following notation makes all this easier to express:

> The **indefinite integral** of a function \( f(x) \) is denoted by

\[
\int f(x) \, dx
\]

and represents the entire family of antiderivatives of \( f(x) \).

The large S-shaped symbol before \( f(x) \) is called an **integral sign**.
Though the indefinite integral \( \int f(x) \, dx \) represents all antiderivatives of \( f(x) \), the integral can be thought of as a single object or function in its own right, whose derivative is \( f'(x) \):

\[
\frac{d}{dx} \left( \int f(x) \, dx \right) = f(x)
\]

You might be wondering what the integral sign in the indefinite integral represents, and why an infinitesimal \( dx \) is included. It has to do with what an infinitesimal represents: an infinitesimal “piece” of a quantity. For an antiderivative \( F(x) \) of a function \( f(x) \), the infinitesimal (or differential) \( dF \) is given by \( dF = F'(x) \, dx = f(x) \, dx \), and so

\[
F(x) = \int f(x) \, dx = \int dF.
\]

The integral sign thus acts as a summation symbol: it sums up the infinitesimal “pieces” \( dF \) of the function \( F(x) \) at each \( x \) so that they add up to the entire function \( F(x) \). Think of it as similar to the usual summation symbol \( \Sigma \) used for discrete sums; the integral sign \( \int \) takes the sum of a continuum of infinitesimal quantities instead.

Finding (or evaluating) the indefinite integral of a function is called integrating the function, and integration is antidifferentiation.

**Example 5.1**

Evaluate \( \int 0 \, dx \).

**Solution:** Since the derivative of any constant function is 0, then \( \int 0 \, dx = C \), where \( C \) is a generic constant.

Note: From now on \( C \) will simply be assumed to represent a generic constant, without having to explicitly say so every time.

**Example 5.2**

Evaluate \( \int 1 \, dx \).

**Solution:** Since the derivative of \( F(x) = x \) is \( F'(x) = 1 \), then \( \int 1 \, dx = x + C \).

**Example 5.3**

Evaluate \( \int x \, dx \).

**Solution:** Since the derivative of \( F(x) = \frac{x^2}{2} \) is \( F'(x) = x \), then \( \int x \, dx = \frac{x^2}{2} + C \).

Since \( \frac{d}{dx} \left( \frac{x^{n+1}}{n+1} \right) = x^n \) for any number \( n \neq -1 \), and \( \frac{d}{dx} (\ln |x|) = \frac{1}{x} = x^{-1} \), then any power of \( x \) can be integrated:
Power Formula: \[ \int x^n \, dx = \begin{cases} 
\frac{x^{n+1}}{n+1} & \text{if } n \neq -1 \\
\ln|x| & \text{if } n = -1 
\end{cases} \]

The following rules for indefinite integrals are immediate consequences of the rules for derivatives:

1. \( \int k f(x) \, dx = k \int f(x) \, dx \)
2. \( \int (f(x) + g(x)) \, dx = \int f(x) \, dx + \int g(x) \, dx \)
3. \( \int (f(x) - g(x)) \, dx = \int f(x) \, dx - \int g(x) \, dx \)

The above rules are easily proved. For example, the first rule is a simple consequence of the Constant Multiple Rule for derivatives: if \( F(x) = \int f(x) \, dx \), then
\[ \frac{d}{dx}(k F(x)) = k \frac{d}{dx}(F(x)) = k f(x) \quad \Rightarrow \quad \int k f(x) \, dx = k F(x) = k \int f(x) \, dx . \]

The other rules are proved similarly and are left as exercises. Repeated use of the above rules along with the Power Formula shows that any polynomial can be integrated term by term—in fact any finite sum of functions can be integrated in that manner:

For any functions \( f_1, \ldots, f_n \) and constants \( k_1, \ldots, k_n \),
\[ \int (k_1 f_1(x) + \cdots + k_n f_n(x)) \, dx = k_1 \int f_1(x) \, dx + \cdots + k_n \int f_n(x) \, dx . \]

**Example 5.4**

Evaluate \( \int (x^7 - 3x^4) \, dx \).

**Solution:** Integrate term by term, pulling constant multiple outside the integral:
\[ \int (x^7 - 3x^4) \, dx = \int x^7 \, dx - 3 \int x^4 \, dx = \frac{x^8}{8} - \frac{3x^5}{5} + C \]
Example 5.5

Evaluate \( \int \sqrt{x} \, dx \).

Solution: Use the Power Formula:

\[
\int \sqrt{x} \, dx = \int x^{1/2} \, dx = \frac{x^{3/2}}{3/2} + C = \frac{2x^{3/2}}{3} + C
\]

Example 5.6

Evaluate \( \int \left( \frac{1}{x^2} + \frac{1}{x} \right) \, dx \).

Solution: Use the Power Formula and integrate term by term:

\[
\int \left( \frac{1}{x^2} + \frac{1}{x} \right) \, dx = \int \left( x^{-2} + x^{-1} \right) \, dx = \frac{x^{-1}}{-1} + \ln|x| + C = -\frac{1}{x} + \ln|x| + C
\]

The following indefinite integrals are just re-statements of the corresponding derivative formulas for the six basic trigonometric functions:

\[
\int \cos x \, dx = \sin x + C
\]
\[
\int \sin x \, dx = -\cos x + C
\]
\[
\int \sec^2 x \, dx = \tan x + C
\]
\[
\int \sec x \tan x \, dx = \sec x + C
\]
\[
\int \csc x \cot x \, dx = -\csc x + C
\]
\[
\int \csc^2 x \, dx = -\cot x + C
\]

Since \( \frac{d}{dx}(e^x) = e^x \), then:

\[
\int e^x \, dx = e^x + C
\]
Example 5.7

Evaluate \( \int (3 \sin x + 4 \cos x - 5e^x) \, dx \).

Solution: Integrate term by term:

\[
\int (3 \sin x + 4 \cos x - 5e^x) \, dx = 3 \int \sin x \, dx + 4 \int \cos x \, dx - 5 \int e^x \, dx
\]

\[
= -3 \cos x + 4 \sin x - 5e^x + C
\]

Example 5.8

Recall from Section 1.1 the example of an object dropped from a height of 100 ft. Show that the height \( s(t) \) of the object \( t \) seconds after being dropped is \( s(t) = -16t^2 + 100 \), measured in feet.

Solution: When the object is dropped at time \( t = 0 \) the only force acting on it is gravity, causing the object to accelerate downward at the known constant rate of 32 ft/s\(^2\). The object’s acceleration \( a(t) \) at time \( t \) is thus \( a(t) = -32 \). If \( v(t) \) is the object’s velocity at time \( t \), then \( v'(t) = a(t) \), which means that

\[
v(t) = \int a(t) \, dt = \int -32 \, dt = -32t + C
\]

for some constant \( C \). The constant \( C \) here is not generic—it has a specific value determined by the initial condition on the velocity: the object was at rest at time \( t = 0 \). That is, \( v(0) = 0 \), which means

\[
0 = v(0) = -32(0) + C \quad \Rightarrow \quad v(t) = -32t
\]

for all \( t \geq 0 \). Likewise, since \( s'(t) = v(t) \) then

\[
s(t) = \int v(t) \, dt = \int -32t \, dt = -16t^2 + C
\]

for some constant \( C \), determined by the initial condition that the object was 100 ft above the ground at time \( t = 0 \). That is, \( s(0) = 100 \), which means

\[
100 = s(0) = -16(0)^2 + C \quad \Rightarrow \quad s(t) = -16t^2 + 100
\]

for all \( t \geq 0 \). √

The formula for \( s(t) \) in Example 5.8 can be generalized as follows: denote the object’s initial position at time \( t = 0 \) by \( s_0 \), let \( v_0 \) be the object’s initial velocity (positive if thrown upward, negative if thrown downward), and let \( g \) represent the (positive) constant acceleration due to gravity. By Newton’s First Law of motion the only acceleration imparted to the object after throwing it is due to gravity:

\[
a(t) = -g \quad \Rightarrow \quad v(t) = \int a(t) \, dt = \int -g \, dt = -gt + C
\]

for some constant \( C \): \( v_0 = v(0) = -g(0)+C = C \). Thus, \( v(t) = -gt+v_0 \) for all \( t \geq 0 \), and so

\[
s(t) = \int v(t) \, dt = \int (-gt + v_0) \, dt = -\frac{1}{2}gt^2 + v_0t + C
\]
for some constant $C$: $s_0 = s(0) = -\frac{1}{2}g(0)^2 + v_0(0) + C$. To summarize:

**Free fall motion:** At time $t \geq 0$:

- **acceleration:** $a(t) = -g$
- **velocity:** $v(t) = -gt + v_0$
- **position:** $s(t) = -\frac{1}{2}gt^2 + v_0t + s_0$
- **initial conditions:** $s_0 = s(0), v_0 = v(0)$

Note that the units are not specified—they just need to be consistent. In metric units, $g = 9.8 \text{ m/s}^2$, while $g = 32 \text{ ft/s}^2$ in English units.

Thinking of an indefinite integral as the sum of all the infinitesimal “pieces” of a function—for the purpose of retrieving that function—provides a handy way of integrating a differential equation to obtain the solution. The key idea is to transform the differential equation into an *equation of differentials*, which has the effect of treating functions as variables. Some examples will illustrate the technique.

**Example 5.9**

For any constant $k$, show that every solution of the differential equation $\frac{dy}{dt} = ky$ is of the form $y = Ae^{kt}$ for some constant $A$. You can assume that $y(t) > 0$ for all $t$.

**Solution:** Put the $y$ terms on the left and the $t$ terms on the right, i.e. separate the variables:

$$\frac{dy}{y} = k \, dt$$

Now integrate both sides (notice how the function $y$ is treated as a variable):

$$\int \frac{dy}{y} = \int k \, dt$$

$$\ln y + C_1 = kt + C_2 \quad (C_1 \text{ and } C_2 \text{ are constants})$$

$$\ln y = kt + C \quad \text{(combine } C_1 \text{ and } C_2 \text{ into the constant } C)$$

$$y = e^{kt+C} = e^{kt} \cdot e^C = A e^{kt}$$

where $A = e^C$ is a constant. Note that this is the formula for radioactive decay from Section 2.3.

**Example 5.10**

Recall from Section 3.6 the equation of differentials

$$\frac{dP}{P} + \frac{dV}{V} = \frac{dT}{T}$$

relating the pressure $P$, volume $V$ and temperature $T$ of an ideal gas. Integrate that equation to obtain the original ideal gas law $PV = RT$, where $R$ is a constant.
Solution: Integrating both sides of the equation yields
\[
\int \frac{dP}{P} + \int \frac{dV}{V} = \int \frac{dT}{T}
\]
\[
\ln P + \ln V = \ln T + C \quad (C \text{ is a constant})
\]
\[
\ln(PV) = \ln T + C
\]
\[
PV = e^{\ln T + C} = e^{\ln T}e^C = Te^C = RT
\]

where \( R = e^C \) is a constant. \( \checkmark \)

The integrals so far have been simple, perhaps giving the impression that integration is easy. The formulas in this section depended on already knowing the derivatives of certain functions and then “working backward” from their derivatives to obtain the original functions. Without that prior knowledge you would be reduced to guessing, or perhaps recognizing a pattern from some derivative you have encountered.

For example, there is no formula yet for \( \int \cos^2 x \, dx \). But by the Chain Rule you know that the derivative of \( \frac{1}{2} \sin 2x \) is \( \cos 2x \), so that \( \int \cos 2x \, dx = \frac{1}{2} \sin 2x + C \).

Recognizing patterns like that will be the subject of later sections. However, in many cases—in fact, \textit{most} cases—no pattern will help. For example, the indefinite integrals \( \int e^{x^2} \, dx \) and \( \int \sin(x^2) \, dx \) cannot be evaluated in a simple closed form. So integration can be not only hard but often impossible.

**Exercises**

**A**

For Exercises 1-15, evaluate the given indefinite integral.

1. \( \int (x^2 + 5x - 3) \, dx \)
2. \( \int 3\cos x \, dx \)
3. \( \int 4e^x \, dx \)
4. \( \int (x^5 - 8x^4 - 3x^3 + 1) \, dx \)
5. \( \int 5\sin x \, dx \)
6. \( \int \frac{3e^x}{5} \, dx \)
7. \( \int \frac{6}{x} \, dx \)
8. \( \int \frac{4}{3x} \, dx \)
9. \( \int (-2\sqrt{x}) \, dx \)
10. \( \int \frac{1}{3\sqrt{x}} \, dx \)
11. \( \int (x + x^{4/3}) \, dx \)
12. \( \int \frac{1}{3\sqrt{x}} \, dx \)
13. \( \int 3\sec x \tan x \, dx \)
14. \( \int 5\sec^2 x \, dx \)
15. \( \int 7\csc^2 x \, dx \)
16. Prove the sum and difference rules for indefinite integrals: \( \int (f(x) \pm g(x)) \, dx = \int f(x) \, dx \pm \int g(x) \, dx \)
17. Integrate both sides of the equation
\[
\frac{dP}{P} + \frac{dM}{M} = \frac{dT}{2T}
\]

...to obtain the ideal gas continuity relation: \( \frac{PM}{\sqrt{T}} = \text{constant} \).
5.2 The Definite Integral

Recall from the last section that the integral sign in the indefinite integral
\[ \int f(x) \, dx \]
represents a summation of the infinitesimals \( f(x) \, dx = dF \) for an antiderivative \( F(x) \) of \( f(x) \). Why is the term “indefinite” used? Because the summation is indefinite: the \( x \) in \( f(x) \, dx \) is defined generically, meaning “\( x \) in general,” that is, not for \( x \) in a specific range of values. The same summation over a specific, definite range of values of \( x \), say, over an interval \([a, b]\), is a different type of integral:

The definite integral of a function \( f(x) \) over an interval \([a, b]\) is denoted by
\[ \int_{a}^{b} f(x) \, dx \]
and represents the sum of the infinitesimals \( f(x) \, dx \) for all \( x \) in \([a, b]\).

An indefinite integral yields a generic function, whereas a definite integral yields either a number or a specific function. There are many ways to calculate the specific summation in a definite integral, one of which is motivated by a geometric interpretation of the infinitesimal \( f(x) \, dx \) as the area of a rectangle, as in Figure 5.2.1 below:

![Figure 5.2.1](image)

The shaded rectangle in the above picture has height \( f(x) \) and width \( dx \), and so its area is \( f(x) \, dx \). In fact, it appears that that area is just a little bit smaller than the area under the curve \( y = f(x) \) and above the \( x \)-axis between \( x \) and \( x + dx \); there is a small gap between the curve and the top of the rectangle, accounting for the difference in the area. However, the area of that gap turns out to be zero, as shown below:
By the Microstraightness Property, the curve $y = f(x)$ shown in Figure 5.2.1 is a straight line over the infinitesimal interval $[x, x + dx]$, as shown in Figure 5.2.2. Thus, the part of the area between the curve and the $x$-axis over the interval $[x, x + dx]$ consists of two parts: the area $f(x) \, dx$ of the shaded rectangle and the area of the right triangle $\triangle ABC$, both of which are shown in Figure 5.2.2. However, the area of $\triangle ABC$ is zero:

\[
\text{Area of } \triangle ABC = \frac{1}{2} \text{(base)} \times \text{(height)} = \frac{1}{2} (dx)(df) = \frac{1}{2} (dx)(f'(x) \, dx) = \frac{1}{2} f'(x)(dx)^2 = 0
\]

The function $f$ shown in Figure 5.2.2 is increasing at $x$, but a similar argument could be made if $f$ were decreasing at $x$. Hence, the area between the curve $y = f(x)$ and the $x$-axis comes solely from the rectangles with area $f(x) \, dx$, as $x$ varies from $a$ to $b$. The sum of all those rectangular areas, though, equals the definite integral of $f(x)$ over $[a, b]$. The definite integral can thus be interpreted as an area:

For a function $f(x) \geq 0$ over $[a, b]$, the area under the curve $y = f(x)$ between $x = a$ and $x = b$, denoted by $A$, is given by

\[
A = \int_{a}^{b} f(x) \, dx
\]

and represents the area of the region $R$ bounded above by $y = f(x)$, bounded below by the $x$-axis, and bounded on the sides by $x = a$ and $x = b$ (with $a < b$).

---

1The function $f$ is assumed to be differentiable at $x$, in this case. If not then the points where $f$ is not differentiable can be excluded without affecting the integral.
The Definite Integral • Section 5.2 133

Figure 5.2.3 The area $A$ of the region $R$ equals $\int_{a}^{b} f(x) \, dx$.

In Figure 5.2.3 the area under the curve $y = f(x)$ between $x = a$ and $x = b$ is the area $A$ of the shaded region $R$, namely $A = \int_{a}^{b} f(x) \, dx$. To calculate that area for a specific function, rectangles can again be used, but this time with widths that are small positive numbers instead of infinitesimals. The procedure is as follows:

1. Create a partition $P = \{x_0 < x_1 < \cdots < x_{n-1} < x_n\}$ of the interval $[a, b]$ into $n \geq 1$ subintervals $[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]$, with $x_0 = a$ and $x_n = b$.

2. In each subinterval $[x_{i-1}, x_i]$ of $P$ pick a number $x^*_i$, so that $x_{i-1} \leq x^*_i \leq x_i$ for $i = 1$ to $n$.

3. For $i = 1$ to $n$, form a rectangle whose base is the subinterval $[x_{i-1}, x_i]$ of length $\Delta x_i = x_i - x_{i-1} > 0$ and whose height is $f(x^*_i)$.

4. Take the sum $f(x^*_1)\Delta x_1 + f(x^*_2)\Delta x_2 + \cdots + f(x^*_n)\Delta x_n$ of the areas of these rectangles, called a Riemann sum.

5. Take the limit of the Riemann sums as $n \to \infty$, so that the subinterval lengths approach 0. If the limit exists then that limit is the area $A$ of the region $R$:

$$\text{Area } A = \int_{a}^{b} f(x) \, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x^*_i)\Delta x_i$$  \hspace{1cm} (5.1)
The limit in Formula (5.1) should be taken over all partitions whose norm—the length of the largest subinterval—approaches 0. In practice, however, the partitions are usually chosen so that the subintervals are of equal length, and then simply make those equal lengths smaller and smaller by dividing the interval \([a, b]\) into more and more such subintervals. Note that the points \(x_i^*\) in each subinterval can be anywhere in the subinterval—often the midpoint of the subinterval is chosen, but the left and right endpoints are also typical choices.

In the above procedure the gaps between the rectangles and the curve will have areas approaching 0 as the number \(n\) of subintervals grows and the subinterval lengths approach 0. This is true if the function \(f\) is differentiable, and in fact even if \(f\) is merely continuous.\(^2\) Thus, the area under the curve can be defined by the above procedure.

To calculate the area under a curve in this manner, the reader should have some familiarity with the summation notation in Formula (5.1).

For real numbers \(a_1, a_2, \ldots, a_n\) and an integer \(n \geq 1\),

\[
\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n
\]

is the sum of \(a_1, \ldots, a_n\). The symbol \(\Sigma\) is called the **summation sign**, which is the Greek capital letter Sigma.

The following rules for this “Sigma notation” are intuitively obvious:

<table>
<thead>
<tr>
<th>Rule</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\sum_{k=1}^{n} (a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k)</td>
</tr>
<tr>
<td>2</td>
<td>(\sum_{k=1}^{n} (a_k - b_k) = \sum_{k=1}^{n} a_k - \sum_{k=1}^{n} b_k)</td>
</tr>
<tr>
<td>3</td>
<td>(\sum_{k=1}^{n} ca_k = c \sum_{k=1}^{n} a_k)</td>
</tr>
<tr>
<td>4</td>
<td>(\sum_{k=1}^{n} a_k = \sum_{i=1}^{n} a_i) (i.e. the sum is independent of the summation index letter)</td>
</tr>
</tbody>
</table>

\(^2\)For a proof and fuller discussion of all this, see Ch.1-2 in KNOPP, M.I., *Theory of Area*, Chicago: Markham Publishing Co., 1969. The book attempts to define precisely what an “area” actually means, including that of a rectangle (showing agreement with the intuitive notion of width times height).
The following summation formulas can be helpful when calculating Riemann sums:

Let $n \geq 1$ be a positive integer. Then:

(1) $\sum_{k=1}^{n} 1 = n$

(2) $\sum_{k=1}^{n} k = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$

(3) $\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$

(4) $\sum_{k=1}^{n} k^3 = 1^3 + 2^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$

(5) $\sum_{k=1}^{n} k^4 = 1^4 + 2^4 + \cdots + n^4 = \frac{n(n+1)(6n^3+9n^2+n-1)}{30}$

Formula (1) is obvious: add the number 1 a total of $n$ times and the sum is $n$. Formula (2) can be proved by induction:

1. Show that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ for $n = 1$:
   $$\sum_{k=1}^{1} k = 1 = \frac{1(1+1)}{2} \checkmark$$

2. Assume that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ for some integer $n \geq 1$. Show that the formula holds for $n$ replaced by $n+1$, that is:
   $$\sum_{k=1}^{n+1} k = \frac{(n+1)((n+1)+1)}{2} = \frac{(n+1)(n+2)}{2}$$
   To show this, note that
   $$\sum_{k=1}^{n+1} k = 1 + 2 + \cdots + n + (n+1) = \sum_{k=1}^{n} k + (n+1)$$
   $$= \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2} \checkmark$$

3. By induction, this proves the formula for all integers $n \geq 1$. QED
Formulas (3)-(5) can be proved similarly by induction (see the exercises). The example below shows how Formulas (2) and (3) are used in finding the limit of a Riemann sum.

**Example 5.11**

Use Riemann sums to calculate \( \int_1^2 x^2 \, dx \).

**Solution:** The definite integral is the area under the curve \( y = f(x) = x^2 \) between \( x = 1 \) and \( x = 2 \), as shown in Figure 5.2.5(a):

\[
\int_1^2 x^2 \, dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*) \Delta x_i = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_{i-1}) \frac{1}{n} = \lim_{n \to \infty} \sum_{i=1}^{n} x_i^{2} \frac{1}{n}
\]

\[
= \lim_{n \to \infty} \sum_{i=1}^{n} \left(1 + \frac{i-1}{n}\right)^2 \frac{1}{n} = \lim_{n \to \infty} \sum_{i=1}^{n} \left(\frac{1}{n} + \frac{2}{n^2}(i-1) + \frac{1}{n^3}(i-1)^2\right)
\]

\[
= \lim_{n \to \infty} \left(\frac{n}{n} + \frac{2}{n^2} \sum_{i=1}^{n} (i-1) + \frac{1}{n^3} \sum_{i=1}^{n} (i-1)^2\right) = \lim_{n \to \infty} \left(1 + \frac{2}{n^2} \sum_{i=1}^{n-1} i + \frac{1}{n^3} \sum_{i=1}^{n-1} i^2\right)
\]

\[
= \lim_{n \to \infty} \left(1 + \frac{2}{n^2} \cdot \frac{(n-1)n}{2} + \frac{1}{n^3} \cdot \frac{(n-1)n(2n-1)}{6}\right) = \lim_{n \to \infty} \left(1 + \frac{2}{n^2} \cdot \frac{n-1}{2} + \frac{1}{n^3} \cdot \frac{n(n-1)}{6}\right)
\]

\[
= \left(\lim_{n \to \infty} 1\right) + \left(\lim_{n \to \infty} \frac{n-1}{n}\right) + \left(\lim_{n \to \infty} \frac{2n^2-3n+1}{6n^2}\right)
\]

\[
= 1 + \frac{1}{1} + \frac{2}{6} = \frac{7}{3}
\]
It is often simpler to use a computer to calculate approximations of a definite integral, by taking the Riemann sum of a sufficiently large number of rectangles in order to achieve the desired accuracy. Choosing subintervals of equal length, as in Example 5.11, makes it easier to use an algorithm to calculate the integral.

For example, the table below summarizes the calculations of Riemann sums for the function in Example 5.11—namely \( f(x) = x^2 \) over \([1, 2]\)—using different values for the points \( x_i^* \) in the subintervals (left endpoints, midpoints, and right endpoints):

<table>
<thead>
<tr>
<th># of rectangles</th>
<th>Left endpoint</th>
<th>Midpoint</th>
<th>Right endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2.25</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1.625</td>
<td>2.3125</td>
<td>3.125</td>
</tr>
<tr>
<td>3</td>
<td>1.851851851852</td>
<td>2.324074074074</td>
<td>2.851851851852</td>
</tr>
<tr>
<td>4</td>
<td>1.96875</td>
<td>2.328125</td>
<td>2.71875</td>
</tr>
<tr>
<td>5</td>
<td>2.04</td>
<td>2.33</td>
<td>2.64</td>
</tr>
<tr>
<td>10</td>
<td>2.185</td>
<td>2.3325</td>
<td>2.485</td>
</tr>
<tr>
<td>100</td>
<td>2.31835</td>
<td>2.33325</td>
<td>2.34835</td>
</tr>
<tr>
<td>1000</td>
<td>2.3318335</td>
<td>2.333325</td>
<td>2.3348335</td>
</tr>
<tr>
<td>10000</td>
<td>2.333183335</td>
<td>2.33333325</td>
<td>2.333483335</td>
</tr>
<tr>
<td>100000</td>
<td>2.3333183335</td>
<td>2.3333333325</td>
<td>2.3333483335</td>
</tr>
<tr>
<td>1000000</td>
<td>2.33333183333</td>
<td>2.3333333333</td>
<td>2.33333483333</td>
</tr>
</tbody>
</table>

Due to the concavity of the curve \( y = x^2 \), using the left endpoints underestimates the actual area, whereas using the right endpoints yields an overestimate. Using the midpoints usually gives better results (i.e. more accuracy in fewer iterations).

So far only definite integrals of nonnegative functions have been considered—that is, functions \( f(x) \geq 0 \) over an interval \([a, b]\). If \( f(x) \) is either negative or changes sign over \([a, b]\), then the definite integral can be defined as follows:

Let \( R \) be the region bounded by \( y = f(x) \) and the \( x \)-axis between \( x = a \) and \( x = b \). If \( f(x) \leq 0 \) over \([a, b]\), then

\[
\int_{a}^{b} f(x) \, dx = \text{the negative of the area of } R
\]

If \( f(x) \) changes sign over \([a, b]\), then

\[
\int_{a}^{b} f(x) \, dx = \text{the net area of } R,
\]

where the parts of \( R \) above the \( x \)-axis count as positive area and the parts below count as negative area.
Note: In the definite integral
\[ \int_{a}^{b} f(x) \, dx \]
the numbers \(a\) and \(b\) are called the **limits of integration**, with \(a\) being the **lower limit of integration** and \(b\) the **upper limit of integration**. The function \(f(x)\) being integrated is called the **integrand**, in both definite and indefinite integrals.

### Exercises

**A**

1. Explain why \(\int_{a}^{b} c \, dx = c(b - a)\) for any constant \(c\).

2. Would using left endpoints in the Riemann sums underestimate or overestimate \(\int_{1}^{2} \ln x \, dx\)? Explain.

**B**

3. Use Riemann sums to calculate \(\int_{0}^{1} x \, dx\).

4. Use Riemann sums to calculate \(\int_{0}^{1} x^2 \, dx\).

5. Use Riemann sums to calculate \(\int_{0}^{1} 3x^2 \, dx\).

6. Use Riemann sums to calculate \(\int_{0}^{1} x^3 \, dx\).

7. Prove the formula \(\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}\) by induction on \(n \geq 1\).

8. Prove the formula \(\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}\) as follows:
   - **(a)** Show that \(\sum_{k=1}^{n} ((k+1)^3 - k^3) = (n+1)^3 - 1\).
   - **(b)** Show that \((k+1)^3 - k^3 = 3k^2 + 3k + 1\).
   - **(c)** Use the formula \(\sum_{k=1}^{n} k = \frac{n(n+1)}{2}\) and parts (a) and (b) to show that \(\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}\).

9. Prove the formula \(\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}\) by induction on \(n \geq 1\).

10. The famous **quicksort algorithm** in computer science is a popular method for placing objects in some order (e.g. numerical, alphabetical). On average the algorithm needs \(O(n \log n)\) comparisons to sort \(n\) objects (here \(\log n\) means the natural logarithm of \(n\)). The proof of that average complexity depends on the inequality
    \[ \sum_{k=2}^{m-1} k \ln k \leq \int_{2}^{m} x \ln x \, dx \]
    for all integers \(m > 2\). Explain why that inequality is true.

**C**

11. Prove the formula \(\sum_{k=1}^{n} k^4 = 1^4 + 2^4 + \cdots + n^4 = \frac{n(n+1)(6n^3 + 9n^2 + n - 1)}{30}\) by induction on \(n \geq 1\).
5.3 The Fundamental Theorem of Calculus

Using Riemann sums to calculate definite integrals can be tedious, as was seen in the previous section. In fact the technique shown in that section depended on the function being a low-degree polynomial, which obviously will not always be the case. Luckily there is a better way, involving antiderivatives, given by the following theorem:

**Fundamental Theorem of Calculus**: Suppose that a function \( f \) is differentiable on \([a, b]\). Then:

**(I)** The function \( A(x) \) defined on \([a, b]\) by

\[
A(x) = \int_a^x f(t) \, dt
\]

is differentiable on \([a, b]\), and

\[
A'(x) = f(x)
\]

for all \( x \) in \([a, b]\).

**(II)** If \( F \) is an antiderivative of \( f \) on \([a, b]\), i.e. \( F'(x) = f(x) \) for all \( x \) in \([a, b]\), then

\[
\int_a^b f(x) \, dx = F(b) - F(a).
\]

The function \( A(x) \) in Part I of the theorem is sometimes called the **area function** because it represents the area under the curve \( y = f(x) \) over the interval \([a, x]\), as shown in Figure 5.3.1 below.

![Figure 5.3.1](image)

The area function \( A(x) = \int_a^x f(t) \, dt \)

To prove Part I, assume that \( f(x) \geq 0 \) on \([a, b]\) as in Figure 5.3.1 (the proofs for \( f(x) \) either negative or switching sign over \([a, b]\) are similar). The goal is to show that for any \( x \) in \([a, b]\) the differential \( dA \) exists and equals \( f(x) \, dx \). First, \( dA = A(x+dx) - A(x) \) is the area under the curve \( y = f(x) \) over the interval \([x, x + dx]\), as shown in Figure 5.3.2 above.

![Figure 5.3.2](image)
By the Microstraightness Property $f$ is a straight line over the infinitesimal interval $[x, x + dx]$, so $f$ must be either increasing, constant, or decreasing over that interval. The three possibilities are shown in Figure 5.3.3:

![Figure 5.3.3](image)

In the case where $f$ is increasing over $[x, x + dx]$, the infinitesimal area $dA$ is the sum of the area of the rectangle of height $f(x)$ and width $dx$ and the area of the right triangle $\triangle ABC$ shown in Figure 5.3.3(a). The area of $\triangle ABC$ is $\frac{1}{2}(df)(dx) = \frac{1}{2}f'(x)(dx)^2 = 0$, so $dA = f(x)dx$.

In the case where $f$ is constant over $[x, x + dx]$, the infinitesimal area $dA$ is the area of the rectangle of height $f(x)$ and width $dx$, as shown in Figure 5.3.3(b). So again, $dA = f(x)dx$.

In the case where $f$ is decreasing over $[x, x + dx]$, the infinitesimal area $dA$ is the sum of the area of the rectangle of height $f(x + dx)$ and width $dx$ and the area of the right triangle $\triangle ABC$ shown in Figure 5.3.3(c). Note that $df < 0$ since $f$ is decreasing, and so the area of $\triangle ABC$ is $\frac{1}{2}(-df)(dx) = -\frac{1}{2}f'(x)(dx)^2 = 0$. Thus,

$$dA = f(x + dx)dx = (f(x + df)dx = f(x)dx + f'(x)(dx)^2 = f(x)dx + 0 = f(x)dx .$$

So in all three cases, $dA = f(x)dx$, and so $A'(x) = \frac{dA}{dx} = f(x)$, which shows that $A(x)$ is differentiable and has derivative $f(x)$. This proves Part I of the Fundamental Theorem of Calculus. ✓

To prove Part II of the theorem, let $F(x)$ be an antiderivative of $f(x)$ over $[a, b]$. Since $A(x) = \int_a^x f(x)dx$ is also an antiderivative of $f(x)$ over $[a, b]$ by Part I of the theorem, then $A(x)$ and $F(x)$ differ by a constant $C$ over $[a, b]$. In other words:
The Fundamental Theorem of Calculus

Section 5.3

By definition $A(a) = 0$, since it is the area under the curve over the interval $[a,a]$ of zero length. Thus,

$$0 = A(a) = F(a) + C \Rightarrow C = -F(a) \Rightarrow A(x) = F(x) - F(a) \text{ for all } x \text{ in } [a, b]$$

and so

$$\int_{a}^{b} f(x) \, dx = A(b) = F(b) - F(a)$$

which proves Part II of the theorem.\(^3\) \checkmark

Note: In some textbooks Part I is called the First Fundamental Theorem of Calculus and Part II is called the Second Fundamental Theorem of Calculus. The following notation provides a shorthand way of writing $F(b) - F(a)$:

$$F(x) \bigg|_{a}^{b} = F(b) - F(a)$$

**Example 5.12**

Calculate $\int_{1}^{2} x^2 \, dx$.

**Solution:** Recall from Example 5.11 in the previous section that the integral equals $7/3$. In that example Riemann sums were used, but Part II of the Fundamental Theorem of Calculus makes the integral much easier to calculate. Since $F(x) = \frac{x^3}{3}$ is an antiderivative of $f(x) = x^2$, then

$$\int_{1}^{2} x^2 \, dx = \left. \frac{x^3}{3} \right|_{1}^{2} = \frac{2^3}{3} - \frac{1}{3} = \frac{7}{3}.$$  

Note in the above example that any antiderivative of $f(x) = x^2$ could have been used, e.g. $F(x) = \frac{x^3}{3} + 5$. Notice that the constant 5 would have been canceled out when evaluating $F(2) - F(1)$. So you do not need to add a generic constant $C$ to the antiderivative of $f(x)$ in a definite integral, as you would in an indefinite integral.

**Example 5.13**

Calculate $\int_{0}^{\pi} \sin x \, dx$.

**Solution:** Since $F(x) = -\cos x$ is an antiderivative of $f(x) = \sin x$, then

$$\int_{0}^{\pi} \sin x \, dx = -\cos x \bigg|_{0}^{\pi} = -\cos \pi - (-\cos 0) = -(-1) - (-1) = 2.$$  

Example 5.14

Calculate $\int_{-1}^{1} x^3 \, dx$.

Solution: Since $F(x) = \frac{x^4}{4}$ is an antiderivative of $f(x) = x^3$, then

$$\int_{-1}^{1} x^3 \, dx = \frac{x^4}{4} \bigg|_{-1}^{1} = \frac{1}{4} - \frac{(-1)^4}{4} = \frac{1}{4} - \frac{1}{4} = 0.$$ 

Example 5.14 is a special case of the following result for odd functions:

If $f$ is an odd function, i.e. $f(-x) = -f(x)$ for all $x$, then

$$\int_{-a}^{a} f(x) \, dx = 0$$

for all $a > 0$ such that $f$ is continuous on $[-a,a]$.

The idea is that since an odd function is symmetric around the origin, then the area between the curve and the $x$-axis over $[0,a]$ will cancel out the area between the curve and the $x$-axis over $[-a,0]$. Both areas are the same but one gets counted as positive and the other negative, as shown in Figure 5.3.4 below:

![Figure 5.3.4 Odd function $f$ over $[-a,a]$](image)

![Figure 5.3.5 Even function $f$ over $[-a,a]$](image)

By symmetry around the $y$-axis, a similar result holds for even functions (see Figure 5.3.5):

If $f$ is an even function, i.e. $f(-x) = f(x)$ for all $x$, then

$$\int_{-a}^{a} f(x) \, dx = 2 \int_{0}^{a} f(x) \, dx$$

for all $a > 0$ such that $f$ is continuous on $[-a,a]$. 
The following rules for definite integrals are a consequence of the corresponding rules for indefinite integrals:

Let \( f \) and \( g \) be continuous functions on \([a, b]\) and let \( k \) be a constant. Then:

1. \[
    \int_{a}^{b} k \, f(x) \, dx = k \int_{a}^{b} f(x) \, dx
\]

2. \[
    \int_{a}^{b} (f(x) + g(x)) \, dx = \int_{a}^{b} f(x) \, dx + \int_{a}^{b} g(x) \, dx
\]

3. \[
    \int_{a}^{b} (f(x) - g(x)) \, dx = \int_{a}^{b} f(x) \, dx - \int_{a}^{b} g(x) \, dx
\]

The following results for definite integrals are a consequence of the Fundamental Theorem of Calculus:

Let \( f \) be a continuous function on \([a, b]\) and suppose that \( a < c < b \). Then:

1. \[
    \int_{a}^{a} f(x) \, dx = 0
\]

2. \[
    \int_{b}^{b} f(x) \, dx = -\int_{a}^{b} f(x) \, dx
\]

3. \[
    \int_{a}^{b} f(x) \, dx = \int_{a}^{c} f(x) \, dx + \int_{c}^{b} f(x) \, dx
\]

For example, if \( F(x) \) is an antiderivative of \( f(x) \) on \([a, b]\), then

\[
    \int_{a}^{c} f(x) \, dx + \int_{c}^{b} f(x) \, dx = (F(c) - F(a)) + (F(b) - F(a)) = F(b) - F(a) = \int_{a}^{b} f(x) \, dx
\]

which proves rule (3).

The following result is a consequence of Part I of the Fundamental Theorem of Calculus along with the Chain Rule:

**Chain Rule for integrals:** Let \( f \) be a continuous function on an interval \( I \) containing \( x = a \), and let \( g(x) \) be a differentiable function on \( I \). If

\[
    F(x) = \int_{a}^{g(x)} f(t) \, dt \quad \text{for all } x \text{ in } I
\]

then \( F'(x) = f(g(x)) \cdot g'(x) \) for all \( x \) in \( I \).
Example 5.15

Let \( F(x) = \int_{0}^{x} e^{-t^2} \, dt \) for all \( x > 0 \). Find \( F'(x) \).

Solution: By the Chain Rule for integrals, with \( f(t) = e^{-t^2} \) and \( g(x) = x^2 \):

\[
F'(x) = f(g(x)) \cdot g'(x) = e^{-(x^2)^2} \cdot (2x) = 2xe^{-x^4}
\]

Exercises

A

For Exercises 1-12, evaluate the given definite integral.

1. \( \int_{0}^{1} x^2 \, dx \)
2. \( \int_{-1}^{1} x^2 \, dx \)
3. \( \int_{0}^{1} x^3 \, dx \)
4. \( \int_{-1}^{1} (x^2 + 3x - 4) \, dx \)
5. \( \int_{1}^{2} \frac{1}{x^2} \, dx \)
6. \( \int_{2}^{3} \frac{1}{x^5} \, dx \)
7. \( \int_{0}^{\pi} \cos x \, dx \)
8. \( \int_{0}^{1} e^x \, dx \)
9. \( \int_{-1}^{1} 2e^x \, dx \)
10. \( \int_{-\pi}^{\pi} \sin x \, dx \)
11. \( \int_{0}^{4} \sqrt{x} \, dx \)
12. \( \int_{-2}^{2} \frac{x^3e^x}{\cos 2x} \, dx \)
13. Show that \( \ln x = \int_{1}^{x} \frac{1}{t} \, dt \) for all \( x > 0 \).
14. Show that \( \int_{a}^{b} f(x) \, dx = -\int_{b}^{a} f(x) \, dx \).
15. Given \( f(x) = \int_{1}^{x} \frac{t}{\sqrt{t^2 + 1}} \, dt \), find \( f'(3) \) and \( f'(-2) \).

B

16. Prove the Chain Rule for integrals.

17. Explain why for any continuous function \( f \) on \([a, b]\), \( \left| \int_{a}^{b} f(x) \, dx \right| \leq \int_{a}^{b} |f(x)| \, dx \).

18. Explain why if \( f(x) \leq g(x) \) on \([a, b]\) then \( \int_{a}^{b} f(x) \, dx \leq \int_{a}^{b} g(x) \, dx \).

C

19. Show that if \( f(x) \) is continuous on \([a, b]\) then there is a number \( c \) in \((a, b)\) such that

\[
\int_{a}^{b} f(x) \, dx = f(c) \cdot (b - a).
\]

20. Let \( f(t) \) be a continuous function for all \( t \geq 0 \), and for each \( x \geq 0 \) define a function \( g(x) \) by

\[
g(x) = \int_{0}^{x} (x-t)f(t) \, dt.
\]

Show that \( g'(x) = \int_{0}^{x} f(t) \, dt \) for all \( x \geq 0 \).
5.4 Integration by Substitution

The integrals encountered so far—whether indefinite or definite—have been the simplest kind, since the antiderivatives were given by known formulas. For example, \( \int \cos x \, dx = \sin x + C \). What if the integral were \( \int \cos 2x \, dx \) instead? No formula has been discussed yet for this integral, and the answer is not \( \sin 2x + C \), since the derivative of \( \sin 2x \) is \( 2\cos 2x \), not \( \cos 2x \). But dividing \( \sin 2x \) by 2 first and then taking the derivative would yield \( \cos 2x \), so that \( \int \cos 2x \, dx = \frac{1}{2} \sin 2x + C \).

Evaluating an integral in such a manner is often done when the function is not too complicated, as the one above. Usually it will not be quite that simple, and so a general technique called substitution is needed. The idea behind substitution is to replace part of the function being integrated by a new variable—typically \( u \)—so that a complicated function of \( x \) is now a simpler function of \( u \) that you know how to integrate.

**Example 5.16**

Evaluate \( \int \cos 2x \, dx \) by substitution.

**Solution:** The \( 2x \) in the cosine function is what makes this integral unknown, so replace it by \( u \): let \( u = 2x \). The integral is now

\[
\int \cos u \, dx
\]

which is a problem because the point of doing substitution is to eliminate all references to the variable \( x \), including in the infinitesimal \( dx \). The entire integral needs to be in terms of \( u \) and \( du \), but the \( dx \) is still there. So put \( dx \) in terms of \( du \):

\[
u = 2x \Rightarrow du = 2dx \Rightarrow dx = \frac{1}{2} du
\]

The integral now becomes

\[
\int \cos u \left( \frac{1}{2} du \right) = \frac{1}{2} \int \cos u \, du = \frac{1}{2} \sin u + C
\]

by the formula already known, just with the letter \( u \) as the variable instead of \( x \). The original integral was in terms of \( x \), so the the final answer—for an indefinite integral—should also be in terms of \( x \). Thus, the final step is to substitute back into the answer what \( u \) equals in terms of \( x \), namely \( 2x \):

\[
\int \cos 2x \, dx = \frac{1}{2} \sin u + C = \frac{1}{2} \sin 2x + C
\]

If the procedure in the above example seems similar to making a substitution when using the Chain Rule to take a derivative, that is because it is similar: you are basically doing the same thing only in reverse. Just as for differentiation, it is not always obvious what part of the function is the best candidate for substitution when performing integration. There is one obvious rule: *never* make the substitution \( u = x \), because that changes nothing.
Example 5.17

Evaluate \( \int e^{-3x} \, dx \).

Solution: The \(-3x\) in the exponential function is what makes this integral unknown, so make the substitution \( u = -3x \), which means that \( du = -3 \, dx \), and so \( dx = \frac{1}{3} \, du \). Thus:

\[
\int e^{-3x} \, dx = \int e^u \left( -\frac{1}{3} \, du \right) = -\frac{1}{3} \int e^u \, du = -\frac{1}{3} e^u + C = -\frac{1}{3} e^{-3x} + C
\]

The above example can be generalized as follows:

\[
\int e^{ax} \, dx = \frac{1}{a} e^{ax} + C \quad \text{for any constant } a \neq 0
\]

Example 5.17 was the special case with \( a = -3 \).

Example 5.18

Evaluate \( \int (1 + 4x)^5 \, dx \).

Solution: You might be tempted to make the substitution \( u = 4x \), but that would then require finding the integral of \((1 + u)^5\), for which there is not yet any formula. But there is a formula for the integral of \( u^5 \). Hence, let \( u = 1 + 4x \), so that \( du = 4 \, dx \Rightarrow dx = \frac{1}{4} \, du \). Thus:

\[
\int (1 + 4x)^5 \, dx = \frac{1}{4} \int u^5 \, du = \frac{1}{4} \frac{u^6}{6} + C = \frac{1}{24} (1 + 4x)^6 + C
\]

Example 5.19

Evaluate \( \int 2x e^{x^2} \, dx \).

Solution: It might be unclear whether you should make the substitution \( u = 2x \) or \( u = x^2 \), but the hint here is that the derivative of the \( x^2 \) inside the exponential function is \( 2x \), which appears outside the exponential function. Indeed, you could check that letting \( u = 2x \) would result in an integral no simpler than the current one (namely, \( \frac{1}{2} \int e^{u^2} \, du \)). So let \( u = x^2 \), which means \( du = 2x \, dx \). Thus:

\[
\int 2x e^{x^2} \, dx = \int e^u \, du = e^u + C = e^{x^2} + C
\]

Example 5.20

Evaluate \( \int x \sqrt{1 + 3x^2} \, dx \).

Solution: Note that the derivative of the \( 1 + 3x^2 \) term inside the square root function is \( 6x \), which is almost the function \( x \) outside the square root—all that is missing is the constant multiple 6. That is a hint to let \( u = 1 + 3x^2 \). Notice also that \( du = 6x \, dx \Rightarrow x \, dx = \frac{1}{6} \, du \), and \( x \, dx \) is the remaining part of the integral outside the square root. Thus:

\[
\int x \sqrt{1 + 3x^2} \, dx = \frac{1}{6} \int \sqrt{u} \, du = \frac{1}{6} \frac{u^{3/2}}{3/2} + C = \frac{1}{9} (1 + 3x^2)^{3/2} + C
\]
Example 5.21

Evaluate \( \int \frac{x^2 \, dx}{\sqrt{x^3 + 9}} \)

Solution: Let \( u = x^3 + 9 \), so that \( du = 3x^2 \, dx \Rightarrow x^2 \, dx = \frac{1}{3} du \). Thus:

\[
\int \frac{x^2 \, dx}{\sqrt{x^3 + 9}} = \int \frac{\frac{1}{3} du}{\sqrt{u}} = \frac{1}{3} \int u^{-\frac{1}{2}} \, du = \frac{1}{3} \frac{u^{1/2}}{1/2} + C = \frac{2}{3} \sqrt{x^3 + 9} + C
\]

Example 5.22

Evaluate \( \int \frac{2x \, dx}{x^2 - 1} \).

Solution: Notice that the numerator \( 2x \) in the function is exactly the derivative of the denominator \( x^2 - 1 \). That is a hint to substitute on the denominator so that the integral is the natural logarithm function. Let \( u = x^2 - 1 \), so that \( du = 2x \, dx \). Thus:

\[
\int \frac{2x \, dx}{x^2 - 1} = \int \frac{du}{u} = \ln|u| + C = \ln|x^2 - 1| + C
\]

Example 5.23

Evaluate \( \int \tan x \, dx \).

Solution: Notice that \( \tan x = \frac{\sin x}{\cos x} \) and that the numerator \( \sin x \) is almost the derivative of the denominator \( \cos x \); all that is missing is a negative sign. That is a hint to substitute on the denominator: let \( u = \cos x \), so that \( du = -\sin x \, dx \Rightarrow \sin x \, dx = -du \). Thus:

\[
\int \tan x \, dx = \int \frac{\sin x}{\cos x} \, dx = \int -\frac{du}{u} = -\ln|u| + C = -\ln|\cos x| + C = \ln|\cos x|^{-1} + C = \ln|\sec x| + C
\]

Example 5.24

Evaluate \( \int \sec x \, dx \).

Solution: Notice that

\[
\int \sec x \, dx = \int \frac{\sec x (\sec x + \tan x)}{\sec x + \tan x} \, dx = \int \frac{\sec x \tan x + \sec^2 x}{\sec x + \tan x} \, dx
\]

and that the numerator in the last integral is the derivative of the denominator: let \( u = \sec x + \tan x \), so that \( du = (\sec x \tan x + \sec^2 x) \, dx \). Thus:

\[
\int \sec x \, dx = \int \frac{du}{u} = \ln|u| + C = \ln|\sec x + \tan x| + C
\]

The following formulas are straightforward consequences of substitution and the derivatives of inverse trigonometric functions discussed in Section 2.2:
For any constant \( a > 0 \):

\[
\int \frac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1}\left(\frac{x}{a}\right) + C \quad \text{(if } |x| < a) \tag{5.2}
\]

\[
\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \tan^{-1}\left(\frac{x}{a}\right) + C \tag{5.3}
\]

\[
\int \frac{dx}{|x|\sqrt{x^2 - a^2}} = \frac{1}{a} \sec^{-1}\left(\frac{x}{a}\right) + C \quad \text{(if } |x| > a) \tag{5.4}
\]

For example, to prove the second formula, recall that \( \frac{d}{dx}(\tan^{-1}x) = \frac{1}{1 + x^2} \). Make the substitution \( u = x/a \), so that \( x = au \) and \( dx = a \, du \). Thus:

\[
\int \frac{dx}{a^2 + x^2} = \int \frac{a \, du}{a^2 + a^2 u^2} = \frac{1}{a} \int \frac{du}{1 + u^2} = \frac{1}{a} \tan^{-1} u + C = \frac{1}{a} \tan^{-1}\left(\frac{x}{a}\right) + C \quad \checkmark
\]

**Example 5.25**

Evaluate \( \int \frac{dx}{\sqrt{4 - 9x^2}} \).

*Solution:* The \( 4 - 9x^2 \) inside the square root is almost of the form \( a^2 - x^2 \), except for the 9. The goal is to have \( u^2 = 9x^2 \), so let \( u = 3x \), which means that \( dx = \frac{1}{3} \, du \) and \( u^2 = 9x^2 \). Thus,

\[
\int \frac{dx}{\sqrt{4 - 9x^2}} = \frac{1}{3} \int \frac{du}{\sqrt{4 - u^2}} = \frac{1}{3} \sin^{-1}\left(\frac{u}{2}\right) + C = \frac{1}{3} \sin^{-1}\left(\frac{3x}{2}\right) + C
\]

by Formula (5.2) with \( a = 2 \).

To use substitution with definite integrals, follow the same procedure as with indefinite integrals but add one extra step: replace the limits of integration \( x = a \) and \( x = b \) in the original integral \( \int_a^b f(x) \, dx \) by \( u = g(a) \) and \( u = g(b) \), respectively, in the new integral involving \( u \), where \( u = g(x) \) is your substitution.

**Example 5.26**

Evaluate \( \int_1^2 (2x + 1)^3 \, dx \).

*Solution:* Let \( u = g(x) = 2x + 1 \), which means that \( dx = \frac{1}{2} \, du \). The upper limit of integration \( x = 2 \) becomes \( u = g(2) = 2(2) + 1 = 5 \) in the new \( u \)-based integral, while the lower limit of integration \( x = 1 \) becomes \( u = g(1) = 2(1) + 1 = 3 \). Thus:

\[
\int_1^2 (2x + 1)^3 \, dx = \frac{1}{2} \int_3^5 u^3 \, du = \frac{1}{8} u^4 \bigg|_3^5 = \frac{1}{8} (5^4 - 3^4) = 68
\]

Note that you could have put everything back in terms of \( x \) at the end, but there was no need to since you would get the same numerical answer.
The following property of definite integrals comes in handy for evaluating certain types of definite integrals:

For any constant \( a \),

\[
\int_0^a f(x) \, dx = \int_0^a f(a - x) \, dx .
\]

This is simple to prove, using the substitution \( u = a - x \), so \( x = a - u \) and \( dx = -du \), while \( x = 0 \) becomes \( u = a \) and \( x = a \) becomes \( u = 0 \) in the limits of integration:

\[
\int_0^a f(x) \, dx = -\int_0^a f(a - u) \, du = \int_0^a f(a - u) \, du = \int_0^a f(a - x) \, dx \quad \checkmark
\]

**Example 5.27**

Evaluate \( \int_0^\pi \frac{x \sin x}{1 + \cos^2 x} \, dx \).

**Solution:** Let \( I = \int_0^\pi \frac{x \sin x}{1 + \cos^2 x} \, dx \). Then by the above property (with \( a = \pi \)):

\[
I = \int_0^\pi \frac{(\pi - x) \sin(\pi - x)}{1 + \cos^2(\pi - x)} \, dx = \int_0^\pi \frac{(\pi - x) \sin x}{1 + \cos^2 x} \, dx = \pi \int_0^\pi \frac{\sin x}{1 + \cos^2 x} \, dx - \int_0^\pi \frac{x \sin x}{1 + \cos^2 x} \, dx
\]

\[
I = \pi \int_0^\pi \frac{\sin x}{1 + \cos^2 x} \, dx - I
\]

\[
2I = \pi \int_0^\pi \frac{\sin x}{1 + \cos^2 x} \, dx = -\pi \tan^{-1}(\cos x) \bigg|_0^\pi = -\pi \left(-\frac{\pi}{4} - \frac{\pi}{4}\right) = \frac{\pi^2}{2}
\]

\[
I = \frac{\pi^2}{4}
\]

**Exercises**

For Exercises 1-24 evaluate the given integral.

1. \( \int (3 \cos 5x + 4 \sin 5x) \, dx \)
2. \( \int e^{2x} + e^{-2x} \, dx \)
3. \( \int \left( x e^{-x^2} + x^2 \cos x^3 \right) \, dx \)
4. \( \int \frac{x - 2}{x^2 - 4x + 9} \, dx \)
5. \( \int \frac{e^x}{1 + e^x} \, dx \)
6. \( \int \frac{1}{1 + e^x} \, dx \)
7. \( \int \sqrt{x^2 + 4} \, dx \)
8. \( \int \cos^2 x \, dx \)
9. \( \int \tan^2 x \, dx \)
10. \( \int \frac{3}{\sqrt{4 - 25x^2}} \, dx \)
11. \( \int \frac{3}{4 + 25x^2} \, dx \)
12. \( \int \sin^2 x \cos^3 x \, dx \)
13. \[ \int_0^1 (2x + 1)^3 \, dx \]
14. \[ \int_0^1 (2x - 1)^3 \, dx \]
15. \[ \int_0^8 x \sqrt{1 + x} \, dx \]
16. \[ \int_0^{\pi/2} 4 \sin(x/2) \, dx \]
17. \[ \int_0^{\pi/4} 4 \sin x \cos x \, dx \]
18. \[ \int_0^{\sqrt{3}} 5x \cos(x^2) \, dx \]
19. \[ \int_8^0 x \alpha \left(\int \frac{x}{x^2 + 2} \, dx \right) \, dx \]
20. \[ \int_0^{\pi/3} e^x \, dx \]
21. \[ \int_1^{\pi/2} \frac{dx}{\sqrt{x(x + 1)}} \]
22. \[ \int_1^{\pi} \frac{x^2 \, dx}{x^2 + 9} \]
23. \[ \int_1^2 \frac{dx}{x^2 - 6x + 9} \]
24. \[ \int_{-3}^1 \frac{x \, dx}{e^{x^2}} \]

25. Evaluate the indefinite integral \[ \int \sin x \cos x \, dx \]

three different ways:
(a) Use the substitution \( u = \sin x \).
(b) Use the substitution \( u = \cos x \).
(c) Use the trigonometric identity \( 2 \sin x \cos x = \sin (2x) \).
(d) Are the three answers from parts (a)-(c) actually different? Explain.

26. For all positive constants \( L \), show the following:
(a) \[ \int_0^L \left(1 - \frac{x}{L}\right)^2 \, dx = \frac{L}{3} \]
(b) \[ \int_{-L/2}^{L/2} \left(1 - \frac{x}{L}\right)^2 \, dx = \frac{L}{3} \]
(c) \[ \int_0^L \left(1 - \frac{x}{L}\right)^3 \left(\frac{x}{L}\right)^2 \, dx = \frac{L}{60} \]

B

27. Recall from trigonometry that \( \sin^2 x = \frac{1}{2} (1 - \cos 2x) \) for all \( x \).
(a) Use the Fundamental Theorem of Calculus to evaluate \[ \int_0^\pi \sin^2 x \, dx . \]
(b) Approximate the integral from part (a) by dividing the interval \([0, \pi]\) into \( n = 2 \) subintervals of equal length, \([0, \pi/2]\) and \([\pi/2, \pi]\), and finding the exact value of the sum of the areas of the rectangles whose heights are determined at the right endpoints of the subintervals.
(c) Repeat part (b) with \( n = 3 \).
(d) Repeat part (b) with \( n = 4 \).
(e) Repeat part (b) with \( n = 6 \).

28. Show that \[ \int \csc x \, dx = -\ln |\csc x + \cot x| + C. \quad (\text{Hint: See Example 5.24.}) \]

C

29. Use the property \[ \int_a^b f(x) \, dx = \int_0^a f(a - x) \, dx \] to show that
\[ \int_0^{\pi/2} \frac{\sin^2 x}{\sin x + \cos x} \, dx = \frac{1}{\sqrt{2}} \ln \left(\sqrt{2} + 1\right) . \]

(Hint: Use Exercise 28 and the sine addition formula.)
30. This exercise is related to Einstein’s famous law $E = mc^2$. The relativistic momentum $p$ of a particle of mass $m$ moving at a speed $v$ along a straight line (say, the $x$-axis) is

$$p = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}} ,$$

where $c$ is the speed of light. The relativistic force on the particle along that line is

$$F = \frac{dp}{dt} ,$$

which is the same formula as Newton’s Second Law of motion in classical mechanics. Assume that the particle starts at rest at position $x_1$ and ends at position $x_2$ along the $x$-axis. The work done by the force $F$ on the particle is:

$$W = \int_{x_1}^{x_2} F \, dx = \int_{x_1}^{x_2} \frac{dp}{dt} \, dx .$$

(a) Show that

$$\frac{dp}{dv} = \frac{m}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} .$$

(b) Use the Chain Rule formula

$$\frac{dp}{dt} = \frac{dp}{dv} \frac{dv}{dx} \frac{dx}{dt}$$

to show that

$$F \, dx = v \frac{dp}{dv} \, dv .$$

(c) Use parts (a) and (b) to show that

$$W = \int_0^v \frac{dp}{dv} \, v \, dv = \int_0^v \frac{mv}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} \, dv .$$

(d) Use part (c) to show that

$$W = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} - mc^2 .$$

(e) Define the relativistic kinetic energy $K$ of the particle to be $K = W$, and define the total energy $E$ to be

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} .$$

So by part (d), $K = E - mc^2$. Show that

$$E^2 = p^2c^2 + (mc^2)^2 .$$

(Hint: Expand the right side of that equation.)

(f) What is $E$ when the particle is at rest?
According to Kepler’s laws of planetary motion, a planet revolving around the Sun follows an elliptical orbit, with the Sun at one of the foci of the ellipse, as in Figure 5.5.1. The distance $d$ between the planet and the Sun varies over the ellipse, reaching a minimum distance and a maximum distance (by the Extreme Value Theorem). How would you find the average distance between the planet and the Sun over one complete orbit? The idea is to generalize the usual notion of an average of numbers.

Recall that for $n$ numbers $x_1, x_2, \ldots, x_n$ the average, denoted by $\bar{x}$, is simply the sum of the numbers divided by how many numbers there are, namely

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$ 

In statistics $\bar{x}$ is called the mean of $x_1, x_2, \ldots, x_n$. This definition makes sense for a finite set of numbers, but in the case of a planet revolving around the Sun, there are an uncountably infinite number of distances between the planet and the Sun, making the above definition impossible to use. A way of taking a sum over an infinite continuum of values is needed instead. Such a method has already been encountered: the definite integral, which is merely a sum of a continuum of infinitesimal quantities.

To motivate the definition of the average value of a function $f$ over a closed interval $[a, b]$, denoted by $\langle f \rangle$, consider a partition

$$P = \{a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b\}$$

that divides $[a, b]$ into $n$ subintervals $[x_{i-1}, x_i]$ of equal length $\Delta x_i = x_i - x_{i-1} = (b - a)/n$, as in Figure 5.5.2. The $n$ function values $f(x_1), f(x_2), \ldots, f(x_n)$ constitute only a finite subset of all the function values $f(x)$ over $[a, b]$, so their average would be an approximation of the true function average $\langle f \rangle$, namely:

$$\langle f \rangle \approx \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n} = \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

By properties of summations, divide the entire sum by the constant $b - a$ and multiply each term in the sum by $b - a$ to get:
Average Value of a Function • Section 5.5 153

\[ \langle f \rangle \approx \frac{1}{b-a} \sum_{i=1}^{n} f(x_i) \cdot \frac{b-a}{n} = \frac{1}{b-a} \sum_{i=1}^{n} f(x_i) \Delta x_i \]

Note that the last summation on the right is just a Riemann sum for the definite integral \( \int_{a}^{b} f(x) \, dx \), with the points \( x_i^* \) chosen to be the right endpoints of the intervals \([x_{i-1}, x_i]\) for \( i = 1 \) to \( n \). Thus, taking the limit of that sum as \( n \to \infty \) (which means including more and more function values in the average) yields the following definition:

The average value \( \langle f \rangle \) of a function \( f \) over a closed interval \([a, b]\) is given by\(^4\)

\[ \langle f \rangle = \frac{1}{b-a} \int_{a}^{b} f(x) \, dx. \]

**Example 5.28**

Find the average value of \( f(x) = x^2 \) over \([0, 1]\).

**Solution:** By definition, with \( a = 0 \) and \( b = 1 \),

\[ \langle f \rangle = \frac{1}{1-0} \int_{0}^{1} f(x) \, dx = \int_{0}^{1} x^2 \, dx = \left. \frac{x^3}{3} \right|_{0}^{1} = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3} \]

Note that this says that if you took all the numbers between 0 and 1 and squared them, then the average of those squares would be 1/3.

**Example 5.29**

Find the average value of \( f(x) = x^2 \) over \([-1, 1]\).

**Solution:** By definition, with \( a = -1 \) and \( b = 1 \),

\[ \langle f \rangle = \frac{1}{1-(-1)} \int_{-1}^{1} f(x) \, dx = \frac{1}{2} \int_{-1}^{1} x^2 \, dx = \left. \frac{x^3}{6} \right|_{-1}^{1} = \frac{1^3}{6} - \frac{(-1)^3}{6} = \frac{1}{3} \]

Note that this is the same as the average over \([0, 1]\), as shown in the previous example. This should make sense, since the function \( f(x) = x^2 \) is symmetric about the y-axis, so the values of \( f(x) \) from \([-1, 0]\) are the same as those from \([0, 1]\). The values from \([-1, 1]\) just duplicate the values from \([0, 1]\) and hence do not change the average.

**Example 5.30**

Find the average value of \( f(x) = \sin x \) over \([0, \pi]\).

**Solution:** By definition, with \( a = 0 \) and \( b = \pi \),

\[ \langle f \rangle = \frac{1}{\pi-0} \int_{0}^{\pi} f(x) \, dx = \frac{1}{\pi} \int_{0}^{\pi} \sin x \, dx = -\frac{1}{\pi} \cos x \bigg|_{0}^{\pi} = -\frac{1}{\pi} (\cos \pi - \cos 0) = -\frac{1}{\pi} (-1 - 1) = \frac{2}{\pi} \]

\(^4\)In some statistics or mathematics texts you might see the notation \( \bar{f} \) for the average value.
Example 5.31

Find the average distance from the ellipse \( \frac{x^2}{25} + \frac{y^2}{9} = 1 \) to the point \((4,0)\).

Solution: Let \(d\) represent the distance from any point \((x,y)\) on the ellipse to the point \((4,0)\), as in Figure 5.5.3. If \((x,y)\) is on the ellipse \(\frac{x^2}{25} + \frac{y^2}{9} = 1\) then \(y^2 = 9(1-x^2) = \frac{9}{25}(25-x^2)\). So by the distance formula, \(d\) is given by

\[
d^2 = (x-4)^2 + (y-0)^2 = (x-4)^2 + y^2 = (x-4)^2 + \frac{9}{25}(25-x^2) = \frac{25(x-4)^2 + 9(25-x^2)}{25} = \frac{16x^2 - 200x + 625}{25}
\]

\[
d^2 = \frac{(4x-25)^2}{25}, \text{ and so taking square roots gives}
\]

\[
d = \pm \frac{4x-25}{5} = \frac{-4x-25}{5} = \frac{25-4x}{5}
\]

for \(-5 \leq x \leq 5\), since \(d = (4x-25)/5 < 0\) on \([-5,5]\) and the distance cannot be negative. Note that by symmetry of the ellipse about the \(x\)-axis, only the upper half of the ellipse is needed for the average distance, since the lower half just duplicates the distances. Hence, the average distance is

\[
\langle d \rangle = \frac{1}{5-(-5)} \int_{-5}^{5} \frac{25-4x}{5} dx = \frac{1}{50} \left(25x - 2x^2\right) \Big|_{-5}^{5} = \frac{1}{50} (125 - 50 - (-125 - 50)) = 5.
\]

Note that the point \((4,0)\) is one of the foci of the ellipse \(\frac{x^2}{25} + \frac{y^2}{9} = 1\) (why?), which as it turns out makes the calculation of the average distance fairly simple.

What if you wanted the average value of a function \(f\) that is not easily integrable? This is actually quite common. One possibility is to use the **Monte Carlo method**. The idea behind it is simple: go back to the usual definition of an average, by taking a very large number \(N\) of random numbers \(x_1, x_2, \ldots, x_N\) in \([a,b]\) and then using the approximation

\[
\langle f \rangle \approx \frac{f(x_1) + f(x_2) + \cdots + f(x_N)}{N}.
\]

This might seem like taking a step backward from calculus, and it is, but it is surprisingly useful, as well as simple to implement with a computer. In addition, it can be shown that as the number of random points in \([a,b]\) increases, the approximations converge to the actual average.
Example 5.32

The Monte Carlo method is easy to implement in Octave, a free open-source programming language and clone of Matlab. Typically only a “one-liner” is needed, owing to Octave’s vectorization—i.e. the ability to perform mathematical operations on entire arrays of objects all at once.

For example, recall from Example 5.28 that the average value of \( f(x) = x^2 \) over \([0,1]\) is \( \frac{1}{3} = 0.33333 \ldots \). Approximate the average value using an array of 100 million (\(10^8\)) random numbers in \([0,1]\):

```
octave> mean(rand(1e8,1).^2)
an = 0.3333292094741531
```

Octave’s `rand(N,k)` command returns \( k \) arrays of \( N \) random numbers between 0 and 1. So `rand(1e8,1)` returns one array of \( 10^8 \) (\(1e8\) in scientific notation) random numbers in \((0,1)\), and the dot in the `rand(1e8,1).^2` command applies the squaring operation (\(^2\)) to each number in that array. The aggregate `mean` function then calculates the array’s mean. Trigonometric, exponential and other functions can be applied to arrays, with the function evaluating each array element individually. In general the command \((b-a)\cdot \text{rand}(N,1)+a\) will return an array of \( N \) random numbers in the interval \((a,b)\).

For example, the function \( f(x) = \sin (x^2) \) cannot be integrated in a closed form, but its average value over \([\pi,2\pi]\) can be approximated easily in Octave (actual average = -0.0415347531416104):

```
octave> mean(sin((pi.*rand(1e8,1)+pi).^2))
an = -0.04153426177596753
```

### Exercises

**A**

For Exercises 1-9, find the average value of the function \( f(x) \) over the given interval.

1. \( f(x) = 1 \), over \([0,3]\)
2. \( f(x) = x \), over \([0,1]\)
3. \( f(x) = x^2 \), over \([0,2]\)
4. \( f(x) = x^3 \), over \([0,2]\)
5. \( f(x) = \sin 2x \), over \([0,\pi/2]\)
6. \( f(x) = e^x \), over \([-1,4]\)
7. \( f(x) = x^3 \), over \([-1,1]\)
8. \( f(x) = \sin x \), over \([-\pi/2,\pi/2]\)
9. \( f(x) = \frac{1}{x} \), over \([1,3]\)
10. Electrical signals are commonly represented by a periodic waveform \( x(t) \), which is a function of time \( t \) and has period \( T \) (i.e. \( T \) is the smallest positive number such that \( x(t+T) = x(t) \) for all \( t \)). The average power of the waveform is defined as the average value of its square over a single period:

\[
\langle x^2(t) \rangle = \frac{1}{T} \int_{0}^{T} x^2(t) \, dt.
\]

(a) Find the average power of the waveform \( x(t) = A \cos(\omega t + \phi) \), where \( A > 0 \) and \( \omega > 0 \) and \( \phi \) are all constants. (Hint: Use the identity \( \cos^2 x = \frac{1+\cos 2x}{2} \).)

(b) The root mean square of a waveform, abbreviated as \( \text{rms} \), is the square root of the average power. Calculate the \( \text{rms} \) of the waveform from part (a). Write your answer in decimal form as a percentage of the amplitude \( A \).
11. An electric circuit with a supplied voltage (electromotive force) \( E \), a capacitor with capacitance \( C \), and a resistor with resistance \( R \), is shown in the picture on the right. When a switch \( s \) in the circuit is opened at time \( t = 0 \) the current \( I \) through the circuit begins to decrease exponentially as a function of time \( t \) (measured in seconds after the switch is opened), given by

\[
I = \frac{E}{R} e^{-\frac{t}{RC}}
\]

for \( t \geq 0 \).

(a) Sketch a rough graph of \( I \) as a function of \( t \).

(b) Note that at time \( t = 0 \) the current is \( I = \frac{E}{R} \) (measured in amperes), which is the familiar formula from Ohm’s Law. That is the peak value of \( I \). What is the current \( I \) at time \( t = 5RC \)? Write your answer in decimal form as a percentage of the peak current (e.g. 0.42 \( \frac{E}{R} \), which would be 42% of the peak current).

(c) Find the average current in the circuit over the time interval \([0, 5RC]\). Write your answer in decimal form as a percentage of the peak current.

12. A spring with spring constant \( k \) and damping constant \( \nu \) connects two point particles with mass \( m \) in a gravitational wave detector. A gravitational wave passes through the detector at time \( t = 0 \) and induces oscillation in the spring, with a period of \( \frac{2\pi}{\Omega} \) and energy \( E \) at time \( t \geq 0 \) given by

\[
E(t) = \frac{1}{4} m R^2 \left( \Omega^2 \sin^2(\Omega t + \phi) + \omega_0^2 \cos^2(\Omega t + \phi) \right),
\]

where \( \omega_0^2 = 2k/m \), \( \phi = \tan^{-1}(2\nu\Omega/(m(\omega_0^2 - \Omega^2))) \), and \( R \) is a constant.

(a) Show that the average energy \( \langle E \rangle \) over one period \([0, 2\pi/\Omega]\) of oscillation is

\[
\langle E \rangle = \frac{1}{8} m R^2 (\omega_0^2 + \Omega^2).
\]

(b) Suppose a large number of identical detectors of this type are uniformly distributed in a planar array at a density of \( \sigma \) detectors per unit area. The energy \( E_\sigma(t) \) imparted to each detector at time \( t \geq 0 \) by a gravitational wave is

\[
E_\sigma(t) = \nu \Omega^2 R^2 \sin^2(\Omega t + \phi).
\]

Show that the average energy \( \langle E_\sigma \rangle \) over one period \([0, 2\pi/\Omega]\) of oscillation is

\[
\langle E_\sigma \rangle = \frac{1}{2} \nu \Omega^2 R^2.
\]

B

13. For the ellipse \( \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \), with \( a > b > 0 \), the foci are the points \((c, 0)\) and \((-c, 0)\), where \( c = \sqrt{a^2 - b^2} \). Find the average distance from the ellipse to either of its foci in terms of the constants \( a \), \( b \), and \( c \).

14. Write a computer program to use the Monte Carlo method with 1 million random points to approximate the average distance from the ellipse \( \frac{x^2}{25} + \frac{y^2}{9} = 1 \) to the point \((0, 0)\) (use symmetry to choose the smallest interval for the points). Could you have used the integral definition instead? Explain.
APPENDIX A

Answers and Hints to Selected Exercises

Chapter 1

Section 1.1 (p. 6)
1. 2t  2. 19.6t  3. −32t + 2  4. 3t^2

Section 1.2 (p. 14)
1. 0  3. 2x + 2  5. −\frac{1}{x+1}  7. −\frac{2}{x^3}
9. \frac{1}{2\sqrt{x+1}}  11. \frac{2x+3}{2\sqrt{x^2+3x+4}}

Section 1.3 (p. 20)

Section 1.4 (p. 26)
1. 2x − 1  3. 4x^5 + \frac{9}{x^7}  5. x \cos x + \sin x
7. \frac{x \cos x + \sin x}{x^2}  9. \frac{2−2t^2}{(1+t^2)^2}  11. \frac{ad−bc}{(cx+d)^2}
13. 2\pi r

Section 1.5 (p. 30)
1. −20(1−5x)^3  3. −\frac{1}{\sqrt{1−2x}}  5. \frac{1−x}{2\sqrt{3(x+1)^2}}
7. −\frac{8(1−t)^3}{(1+t)^3}  9. 2\sin x \cos x  11. 15\sec^2(5x)
13. 2x \sec(x^2) \tan(x^2)  15. \beta(1−\beta^2)^{-3/2}

Section 1.6 (p. 35)
1. 6x + 2  3. −9\cos 3x  5. \frac{2}{x^3}

Chapter 2

Section 2.1 (p. 40)
1. \(f^{-1}(x) = x, (f^{-1})'(x) = 1\)
3. \(f^{-1}(x) = \sqrt{x}, (f^{-1})'(x) = \frac{1}{2\sqrt{x}}\)
5. \(f^{-1}(x) = \frac{1}{x}, (f^{-1})'(x) = −\frac{1}{x^2}\)
7. \(f^{-1}(x) = \frac{1}{\sqrt{x}}, (f^{-1})'(x) = −\frac{1}{2}x^{-3/2}\)

Section 2.2 (p. 44)
1. 6\sec^2 3x \tan 3x  3. −3\csc^2 3x  5. \frac{3}{9+x^2}
7. −\frac{3}{1+9x^2}  9. \frac{1}{1+x^2}  11. \frac{6\sin^{-1}3x}{\sqrt{1−9x^2}}
13. \frac{1}{1+x^2}  15. \cot^{-1}x − \frac{x}{1+x^2}
Section 2.3 (p. 52)

1. $2e^{2x}$  
2. $e^{x^2}$  
3. $-e^{-x} - e^x$  
4. $\frac{2e^x}{(1-e^x)^2}$  
5. $e^{ex}$  
6. $x^2(x + 2x \ln x)$  
7. $x^2 \left(\cos x \ln x + \frac{\sin x}{x}\right)$  
8. $15.5$ hours  
9. $\frac{1}{x}$  
10. $\frac{6x(\ln(x^2))^2 \sec^2 x^2}{\tan x^2}$  
11. $\frac{2x}{(\ln(2)x^2+1)}$  
12. $\cos(\log_2 x) \cdot \frac{x}{\ln 2}$  
13. $3x^2$

Section 2.4 (p. 55)

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

Section 3.1 (p. 61)

1. $y = 4x - 3$  
2. $y = -6x + 10$  
3. $y = 4x$  
4. $y = x + 3$  
5. $y = 2x$  
6. $y = 240x + 176$  
7. $y = 3x + 31/27$  
8. $y = 3x + 1$  
9. $y = 75.96\degree$  
10. $0\degree$  
11. $90\degree$  
12. $116.6\degree$  
13. $5.71\degree$  
14. $y = -\frac{1}{4}x^2 + \frac{11}{2}$  
15. $y = -\frac{1}{4}x^2 - \frac{81}{4}$, $y = -\frac{1}{4}x + \frac{1159}{108}$

Section 3.2 (p. 72)

1. $\frac{7}{3}$  
2. $0$  
3. $5$  
4. $-1$  
5. $0$  
6. $2$  
7. $11$  
8. $0$  
9. $14$  
10. $\frac{1}{2}$  
11. $15$  
12. $0$  
13. $17$

Section 3.3 (p. 78)

1. continuous  
2. discontinuous  
3. continuous  
4. discontinuous  
5. discontinuous  
6. continuous  
7. discontinuous  
8. continuous  
9. continuous  
10. discontinuous  
11. continuous  
12. continuous  
13. discontinuous  
14. $19$  
15. $e^{-1}$  

Section 3.4 (p. 81)

1. $\frac{-3x^2y+4y^2+2x}{x^2-8xy-1}$  
2. $\frac{2x-8y+1+3(3+2)}{2x-8y+1+3(3+2)}$  
3. $\frac{2y+1}{x+2y}$  
4. $\frac{2}{x^2-1}$  
5. $\frac{2x+1}{x+2y}$  
6. $\frac{2}{x+2y}$  
7. $\frac{2x+1}{x+2y}$  
8. $\frac{2x+1}{x+2y}$  
9. $\frac{2x+1}{x+2y}$  
10. $\frac{2x+1}{x+2y}$  
11. $\frac{2x+1}{x+2y}$  
12. $\frac{2x+1}{x+2y}$  
13. $\frac{2x+1}{x+2y}$

Section 3.5 (p. 83)

1. $80\pi$ ft/s  
2. $2.4$ ft/s  
3. $10$ ft/s  
4. $-76\pi$ cm$^3$/min  
5. $9.4514$ mph  
6. $155.8$ ft/min

Section 3.6 (p. 88)

1. $(2x - 2)dx$  
2. $4x \sin(x^2) \cos(x^2)dx$  
3. $4x \sin(x^2) \cos(x^2)dx$  
4. $11$. Hint: Mimic Example 3.35.

Section 4.1 (p. 99)

1. $125,000$ sq yd  
2. $\frac{V}{2}$  
3. $R = r$  
4. $2ab$  
5. $(1,1)$  
6. $10\sqrt{3}$  
7. $12.8$ ft  
8. $x = \sqrt{\frac{R}{N}}$  
9. $\cos(x) e^{\sin(y)} + x^2 y^3 e^{\sin(x^2)} e^{\cos(x)}$  
10. $\frac{x^3}{3} + \frac{y^3}{2}$  
11. $\frac{x^{3/2} + y^{3/2}}{3}$  
12. $\frac{a^{2/3} + b^{2/3}}{3}$  
13. Hint: You can leave your answer in terms of $R$ and an angle satisfying a certain equation.

Section 4.2 (p. 107)

2. local maximum at $x = 0$, local minimum at $x = 2$, inflection pt at $x = 1$, increasing for $x < 0$ and $x < 2$, decreasing for $0 < x < 2$, concave up for $x > 1$, concave down for $x < 1$  
3. local maximum at $x = 1$, inflection pt at $x = 2$, increasing for $x < 1$, decreasing for $x > 1$, concave up for $x > 2$, concave down for $x < 2$, horizontal asymptote: $y = 0$  
4. $5$. local maximum at
Appendix A: Answers and Hints to Selected Exercises

1. \( x = 0 \), inflection pts at \( x = \pm \frac{1}{\sqrt{3}} \), increasing for \( x < 0 \), decreasing for \( x > 0 \), concave up for \( x < -\frac{1}{\sqrt{3}} \) and \( x > \frac{1}{\sqrt{3}} \), concave down for \(-\frac{1}{\sqrt{3}} < x < \frac{1}{\sqrt{3}} \), horizontal asymptote: \( y = 0 \)

7. local maximum at \( x = \ln 2 \), inflection pt at \( x = \ln 4 \), increasing for \( x < \ln 2 \), decreasing for \( x > \ln 2 \), concave up for \( x > \ln 4 \), concave down for \( x < \ln 4 \), horizontal asymptote: \( y = 0 \)

Section 4.3 (p. 116)
1. \( x = 0.450184 \) 3. \( x = 0.567143 \) 5. \( x = 1.414213 \) 11. global maximum at \( x = 2.8214 \)

Section 4.4 (p. 121)
1. No 3. Yes 6. No 18. Hint: Calculate \( f'(x) \) and use the cosine addition formula.

Chapter 5

Section 5.1 (p. 130)
1. \( \frac{x^3}{3} + \frac{5x^2}{2} - 3x + C \) 3. \( 4e^x + C \) 5. \( -5\cos x + C \) 7. \( 6\ln|x| + C \) 9. \( -\frac{4}{3}x^{3/2} + C \) 11. \( \frac{x^2}{2} + \frac{3}{7}x^{7/3} + C \) 13. \( 3\sec x + C \) 15. \( -7\cot x + C \)

Section 5.2 (p. 138)
3. \( \frac{1}{2} \) 4. \( \frac{1}{3} \) 5. 1 6. \( \frac{1}{4} \)

Section 5.3 (p. 144)
1. \( \frac{1}{3} \) 3. \( \frac{1}{4} \) 5. \( \frac{1}{2} \) 7. 1 9. \( 2e - 2e^{-1} \) 11. \( \frac{16}{3} \)

Section 5.4 (p. 149)
1. \( \frac{3\sin 5x - 4\cos 5x}{5} + C \) 3. \( -\frac{1}{2}e^{-\frac{x^2}{2}} + \frac{1}{3}\sin x^3 + C \) 5. \( \ln(1 + e^x) + C \) 7. \( \frac{2}{3}(x + 4)^{5/2} - \frac{8}{3}(x + 4)^{3/2} + C \) 9. \( \tan x - x + C \) 11. \( \frac{3}{10}\tan^{-1}\left(\frac{5x}{2}\right) + C \) 13. 10 15. \( \frac{1192}{15} \) 17. 1 19. \( -\frac{1}{48} \) 21. \( \frac{\pi}{6} \) 23. \( \frac{1}{2} \)

Section 5.5 (p. 155)
1. 1 3. \( \frac{4}{3} \) 5. \( \frac{2}{\pi} \) 7. 0 9. \( \frac{1}{2}\ln 3 \)
GNU Free Documentation License

Version 1.3, 3 November 2008

<http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”,


below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “**you**”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A **“Modified Version”** of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A **“Secondary Section”** is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The **“Invariant Sections”** are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The **“Cover Texts”** are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A **“Transparent”** copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called **“Opaque”**.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or
PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “publisher” means any person or entity that distributes copies of the Document to the public.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgments”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in
addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.
E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document’s license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled “History”, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled “History” in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the “History” section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled “Acknowledgments” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgments and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their
titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgments”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you
follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgments”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.
However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

“Massive Multiauthor Collaboration Site” (or “MMC Site”) means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A “Massive Multiauthor Collaboration” (or “MMC”) contained in the site means any set of copyrightable works thus published on the MMC site.

“CC-BY-SA” means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.
“Incorporate” means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is “eligible for relicensing” if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with . . . Texts.” line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.
This section contains the revision history of the book. For persons making modifications to the book, please record the pertinent information here, following the format in the first item below.

1. VERSION: 0.1
   Date: 2016-01-24
   Author(s): Michael Corral
   Title: Elementary Calculus
   Modification(s): Initial version

2. VERSION: 0.5
   Date: 2020-05-27
   Author(s): Michael Corral
   Title: Elementary Calculus
   Modification(s): Numerous corrections of typos and errors, as well as enhancements of some sections and more exercises.
## Index

### Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta x$</td>
<td>8</td>
</tr>
<tr>
<td>$\langle f \rangle$</td>
<td>152</td>
</tr>
<tr>
<td>$f$</td>
<td>15</td>
</tr>
<tr>
<td>$dx$</td>
<td>15</td>
</tr>
<tr>
<td>$d^2y$</td>
<td>32</td>
</tr>
<tr>
<td>$d^2z$</td>
<td>32</td>
</tr>
<tr>
<td>$d^r\pi$</td>
<td>46</td>
</tr>
<tr>
<td>$\gg$</td>
<td>46</td>
</tr>
<tr>
<td>$\ll$</td>
<td>19</td>
</tr>
<tr>
<td>$I$</td>
<td>4</td>
</tr>
<tr>
<td>$N$</td>
<td>4</td>
</tr>
<tr>
<td>$Q$</td>
<td>4</td>
</tr>
<tr>
<td>$R$</td>
<td>4</td>
</tr>
<tr>
<td>$f'$</td>
<td>8</td>
</tr>
<tr>
<td>$f^{(n)}$</td>
<td>32</td>
</tr>
</tbody>
</table>

### A

<table>
<thead>
<tr>
<th>Concept</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolute value</td>
<td>13</td>
</tr>
<tr>
<td>accelerating</td>
<td>34</td>
</tr>
<tr>
<td>acceleration</td>
<td>33</td>
</tr>
<tr>
<td>algebraic curve</td>
<td>80</td>
</tr>
<tr>
<td>angle of tangent line</td>
<td>59</td>
</tr>
<tr>
<td>antiderivative</td>
<td>123</td>
</tr>
<tr>
<td>arc</td>
<td>18</td>
</tr>
<tr>
<td>area function</td>
<td>139</td>
</tr>
<tr>
<td>area under a curve</td>
<td>132</td>
</tr>
<tr>
<td>asymptote</td>
<td></td>
</tr>
<tr>
<td>horizontal</td>
<td>67</td>
</tr>
<tr>
<td>vertical</td>
<td>66</td>
</tr>
<tr>
<td>average value</td>
<td>152</td>
</tr>
</tbody>
</table>

### B

<table>
<thead>
<tr>
<th>Concept</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>base $a$ logarithm</td>
<td>54</td>
</tr>
<tr>
<td>Big O notation</td>
<td>68</td>
</tr>
<tr>
<td>bisection method</td>
<td>108</td>
</tr>
<tr>
<td>Bohr radius</td>
<td>106</td>
</tr>
</tbody>
</table>

### C

<table>
<thead>
<tr>
<th>Concept</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ceiling function</td>
<td>74</td>
</tr>
<tr>
<td>Chain Rule</td>
<td>27</td>
</tr>
<tr>
<td>Chebyshev polynomials</td>
<td>44</td>
</tr>
<tr>
<td>closed interval</td>
<td>91</td>
</tr>
<tr>
<td>concave down</td>
<td>101</td>
</tr>
<tr>
<td>concave up</td>
<td>101</td>
</tr>
<tr>
<td>concavity</td>
<td>101</td>
</tr>
<tr>
<td>constant function</td>
<td>9</td>
</tr>
<tr>
<td>Constant Multiple Rule</td>
<td>21</td>
</tr>
<tr>
<td>continuity</td>
<td>73</td>
</tr>
<tr>
<td>continuous function</td>
<td>73</td>
</tr>
<tr>
<td>countable</td>
<td>7</td>
</tr>
<tr>
<td>crackle</td>
<td>34</td>
</tr>
<tr>
<td>critical point</td>
<td>91</td>
</tr>
</tbody>
</table>

### D

<table>
<thead>
<tr>
<th>Concept</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darboux's Theorem</td>
<td>121</td>
</tr>
<tr>
<td>decelerating</td>
<td>34</td>
</tr>
<tr>
<td>definite integral</td>
<td>131</td>
</tr>
<tr>
<td>derivative</td>
<td>3, 8</td>
</tr>
<tr>
<td>$n$-th</td>
<td>32</td>
</tr>
<tr>
<td>Chain Rule</td>
<td>27</td>
</tr>
<tr>
<td>Constant Multiple Rule</td>
<td>21</td>
</tr>
<tr>
<td>Difference Rule</td>
<td>21</td>
</tr>
<tr>
<td>exponential function</td>
<td>46</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>first</td>
<td>32</td>
</tr>
<tr>
<td>higher order</td>
<td>32</td>
</tr>
<tr>
<td>natural logarithm function</td>
<td>48</td>
</tr>
<tr>
<td>Power Rule</td>
<td>24</td>
</tr>
<tr>
<td>Product Rule</td>
<td>21</td>
</tr>
<tr>
<td>Quotient Rule</td>
<td>21</td>
</tr>
<tr>
<td>second</td>
<td>32</td>
</tr>
<tr>
<td>Sum Rule</td>
<td>21</td>
</tr>
<tr>
<td>trigonometric functions</td>
<td>23</td>
</tr>
<tr>
<td>with respect to $x$</td>
<td>15</td>
</tr>
<tr>
<td>zero-th</td>
<td>34</td>
</tr>
<tr>
<td>Difference Rule</td>
<td>21</td>
</tr>
<tr>
<td>differentiable</td>
<td>14</td>
</tr>
<tr>
<td>differential</td>
<td>84</td>
</tr>
<tr>
<td>differential equation</td>
<td>36, 47</td>
</tr>
<tr>
<td>differentiation</td>
<td>14</td>
</tr>
<tr>
<td>implicit</td>
<td>79</td>
</tr>
<tr>
<td>discontinuous</td>
<td>73</td>
</tr>
<tr>
<td>domain</td>
<td>8, 37</td>
</tr>
<tr>
<td>dot notation</td>
<td>15</td>
</tr>
</tbody>
</table>

**E**
- $e$                                                45
- eigenfunction                                     36
- eigenvalue                                         36
- elasticity                                          88
- elliptic curve                                     79
- enthalpy                                            89
- entropy                                            89
- equation                                            
  - differential                                      36, 47
  - Schrödinger                                       36
- Eudoxus                                             6
- Euler number                                        45
- even function                                       13
- exponential decay                                  50
- exponential functions                              45
- exponential growth                                 51
- Extended Mean Value Theorem                         120

**F**
- factorial                                           34
- Fermat's Principle                                  95, 100
- First Derivative Test                               104
- floor function                                       74
- free fall formulas                                  128, 129
- function                                             37
  - absolute value                                     13
  - area                                                139
  - average value                                      152
  - ceiling                                             74
  - composition                                         27
  - constant                                            9
  - continuous                                          73
  - differentiable                                      14
  - even                                                13
  - exponential                                         45
  - floor                                               74
  - infinitesimal                                       72
  - inverse                                             38
  - inverse trigonometric                               41
  - linear                                              9
  - natural logarithm                                   47
  - odd                                                 13, 142
  - rational                                            70
  - root                                                108
  - sigmoid neuron                                      52
  - step                                                74
  - trigonometric                                       41
- Fundamental Theorem of Calculus                     139

**G**
- Gibbs energy                                         89

**H**
- half-life                                             50
- horizontal asymptote                                 67

**I**
- ideal gas                                            84
- implicit differentiation                             79
- indefinite integral                                  124
- indeterminate form                                   69, 75
- induction                                            24
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>infinite series</td>
<td>4</td>
</tr>
<tr>
<td>infinitesimal</td>
<td>15</td>
</tr>
<tr>
<td>infinitesimal function</td>
<td>72</td>
</tr>
<tr>
<td>inflection point</td>
<td>102</td>
</tr>
<tr>
<td>instantaneous rate of change</td>
<td>3, 8</td>
</tr>
<tr>
<td>integers</td>
<td>4</td>
</tr>
<tr>
<td>integral</td>
<td>4</td>
</tr>
<tr>
<td>definite</td>
<td>131</td>
</tr>
<tr>
<td>indefinite</td>
<td>124</td>
</tr>
<tr>
<td>integral sign</td>
<td>124</td>
</tr>
<tr>
<td>integrand</td>
<td>138</td>
</tr>
<tr>
<td>integration</td>
<td>125</td>
</tr>
<tr>
<td>inverse function</td>
<td>38</td>
</tr>
<tr>
<td>inverse trigonometric functions</td>
<td>41</td>
</tr>
<tr>
<td>irrational numbers</td>
<td>5</td>
</tr>
<tr>
<td>jerk</td>
<td>34</td>
</tr>
<tr>
<td>jump discontinuity</td>
<td>73</td>
</tr>
<tr>
<td>left continuous</td>
<td>73</td>
</tr>
<tr>
<td>left continuous</td>
<td>73</td>
</tr>
<tr>
<td>Leibniz notation</td>
<td>15</td>
</tr>
<tr>
<td>limit</td>
<td>3, 62</td>
</tr>
<tr>
<td>infinite</td>
<td>66</td>
</tr>
<tr>
<td>left</td>
<td>65</td>
</tr>
<tr>
<td>one-sided</td>
<td>64</td>
</tr>
<tr>
<td>right</td>
<td>65</td>
</tr>
<tr>
<td>limits of integration</td>
<td>138</td>
</tr>
<tr>
<td>line</td>
<td></td>
</tr>
<tr>
<td>normal</td>
<td>60</td>
</tr>
<tr>
<td>tangent</td>
<td>56</td>
</tr>
<tr>
<td>linear combination</td>
<td>24</td>
</tr>
<tr>
<td>linear function</td>
<td>9</td>
</tr>
<tr>
<td>linear operator</td>
<td>24</td>
</tr>
<tr>
<td>logarithmic differentiation</td>
<td>49, 86</td>
</tr>
<tr>
<td>Lorentz factor</td>
<td>72</td>
</tr>
<tr>
<td>Mach number</td>
<td>89</td>
</tr>
<tr>
<td>magnitude</td>
<td>34</td>
</tr>
<tr>
<td>maximum</td>
<td>90</td>
</tr>
<tr>
<td>Mean Value Theorem</td>
<td>117</td>
</tr>
<tr>
<td>alternative form</td>
<td>120</td>
</tr>
<tr>
<td>Extended</td>
<td>120</td>
</tr>
<tr>
<td>method of exhaustion</td>
<td>6</td>
</tr>
<tr>
<td>Microstraightness Property</td>
<td>17</td>
</tr>
<tr>
<td>minimum</td>
<td>90</td>
</tr>
<tr>
<td>Monte Carlo method</td>
<td>154</td>
</tr>
<tr>
<td>natural logarithm</td>
<td>47</td>
</tr>
<tr>
<td>natural numbers</td>
<td>4</td>
</tr>
<tr>
<td>Newton's method</td>
<td>108</td>
</tr>
<tr>
<td>norm of a partition</td>
<td>134</td>
</tr>
<tr>
<td>normal line</td>
<td>60</td>
</tr>
<tr>
<td>N&lt;sup&gt;th&lt;/sup&gt; Derivative Test</td>
<td>105</td>
</tr>
<tr>
<td>O notation</td>
<td>68</td>
</tr>
<tr>
<td>odd function</td>
<td>13, 142</td>
</tr>
<tr>
<td>one-to-one</td>
<td>37</td>
</tr>
<tr>
<td>open interval</td>
<td>91</td>
</tr>
<tr>
<td>optimization</td>
<td>90</td>
</tr>
<tr>
<td>partition</td>
<td>133</td>
</tr>
<tr>
<td>norm</td>
<td>134</td>
</tr>
<tr>
<td>periodic waveform</td>
<td>155</td>
</tr>
<tr>
<td>point of tangency</td>
<td>56</td>
</tr>
<tr>
<td>pop</td>
<td>34</td>
</tr>
<tr>
<td>Power Formula</td>
<td>126</td>
</tr>
<tr>
<td>Power Rule</td>
<td>24</td>
</tr>
<tr>
<td>power series</td>
<td>4</td>
</tr>
<tr>
<td>Product Rule</td>
<td>21</td>
</tr>
<tr>
<td>proof by induction</td>
<td>24</td>
</tr>
<tr>
<td>q-derivative</td>
<td>89</td>
</tr>
<tr>
<td>q-differential</td>
<td>89</td>
</tr>
<tr>
<td>quantum calculus</td>
<td>89</td>
</tr>
<tr>
<td>quicksort algorithm</td>
<td>138</td>
</tr>
<tr>
<td>Quotient Rule</td>
<td>21</td>
</tr>
</tbody>
</table>
Index 173

R
radioactive decay .......................... 50
range .......................... 37
rational function ......................... 70
rational numbers ......................... 4
real numbers .......................... 4
regular polygon ......................... 6
related rates ......................... 82
relativistic momentum ................... 151
relativistic velocity addition law ....... 122
Riemann sum ......................... 133
right continuous ....................... 73
Rolle’s Theorem ......................... 118
root mean square ....................... 155
root of function ....................... 108
rotational velocity ..................... 31

S
Schrödinger equation ..................... 36
secant line .......................... 58
secant method ......................... 111
Second Derivative Test ................. 91
sequence .......................... 4
series .......................... 4
sigma notation ......................... 134
sigmoid neuron function ............... 52
slope .......................... 9
slope of a curve ....................... 58
smooth curve ....................... 18
snap .......................... 34
Snell’s Law .......................... 100
speed .......................... 2, 34
Squeeze Theorem ......................... 71
stationary point ......................... 91
step function .......................... 74
straight line motion ................... 1, 33, 128
substitution method .................. 145
Sum Rule .......................... 21

T
tangent line .......................... 56
angle of .......................... 59
Thales’ Theorem .......................... 18
time dilation ......................... 72
trigonometric functions ............... 23, 41

V
velocity .......................... 2
average .......................... 2
instantaneous .......................... 2
vertical asymptote ................... 66

W
work .......................... 151