Concepts in Calculus I
Second Edition

Miklós Bóna and Sergei Shabanov
University of Florida Department of Mathematics
# Contents

## Chapter 1. Functions
1. Functions 1
2. Classes of Functions 4
3. Operations on Functions 8
4. Viewing the Graphs of Functions 12
5. Inverse Functions 15
6. The Velocity Problem and the Tangent Problem 21

## Chapter 2. Limits and Derivatives
7. The Limit of a Function 25
8. Limit Laws 33
9. Continuous Functions 39
10. Limits at Infinity 43
11. Derivatives 48
12. The Derivative as a Function 51

## Chapter 3. Rules of Differentiation
13. Derivatives of Polynomial and Exponential Functions 57
14. The Product and Quotient Rules 61
15. Derivatives of Trigonometric Functions 64
16. The Chain Rule 67
17. Implicit Differentiation 71
18. Derivatives of Logarithmic Functions 74
19. Applications of Rates of Change 77
20. Related Rates 82
21. Linear Approximations and Differentials 89

## Chapter 4. Applications of Differentiation
22. Minimum and Maximum Values 99
23. The Mean Value Theorem 106
24. The First and Second Derivative Tests 116
25. Taylor Polynomials and the Local Behavior of a Function 125
26. L’Hospital’s Rule 133
27. Analyzing the Shape of a Graph 140
28. Optimization Problems 146
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.</td>
<td>Newton’s Method</td>
<td>153</td>
</tr>
<tr>
<td>30.</td>
<td>Antiderivatives</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>Chapter 5. <strong>Integration</strong></td>
<td></td>
</tr>
<tr>
<td>31.</td>
<td>Areas and Distances</td>
<td>167</td>
</tr>
<tr>
<td>32.</td>
<td>The Definite Integral</td>
<td>176</td>
</tr>
<tr>
<td>33.</td>
<td>The Fundamental Theorem of Calculus</td>
<td>188</td>
</tr>
<tr>
<td>34.</td>
<td>Indefinite Integrals and the Net Change</td>
<td>194</td>
</tr>
<tr>
<td>35.</td>
<td>The Substitution Rule</td>
<td>200</td>
</tr>
</tbody>
</table>
1. Functions

A function $f$ is a rule that associates to each element $x$ in a set $D$ a unique element $f(x)$ of another set $R$. Here the set $D$ is called the domain of $f$, while the set $R$ is called the range of $f$. The fact that $f$ associates to each element of $D$ an element of $R$ is represented by the symbol $f : D \rightarrow R$. Instead of saying that $f$ associates $f(x)$ to $x$, we often say that $f$ sends $x$ to $f(x)$, which is shorter. See Figure 1.1 for an illustration.

![Figure 1.1. Domain and range.](image)

If the sets mentioned in the previous definition are sets of numbers, then it is often easier to describe $f$ by an algebraic expression. Let $\mathbb{N}$ be the set of all natural numbers (which are the nonnegative integers). Then the function $f : \mathbb{N} \rightarrow \mathbb{N}$ given by the rule $f(x) = 2x + 3$ is the function that sends each nonnegative integer $n$ to the nonnegative integer $2n + 3$. For instance, it sends 0 to 3, 1 to 5, 17 to 37, and so on. In this case, the algebraic description is simpler than actually saying “$f$ is the function that sends $n$ to $2n + 3$.”

The rule that describes $f$ may be simple or complicated. It could be that a function is defined by cases such as

$$
f(x) = \begin{cases} 
0.1x & \text{if } 0 \leq x \leq 40, \\
4 + 0.15(x - 40) & \text{if } 40 < x \leq 80, \\
10 + 0.2(x - 80) & \text{if } x > 80.
\end{cases}
$$

This example could describe an income tax code. The first $40,000$ of income is taxed at a rate of 10%, income above $40,000$ but below $60,000$ is taxed at a rate of 15%, and income above $80,000$ is
taxed at a rate of 20%. The value of \( f(x) \) is the amount of tax to be paid after an income of \( x \) thousand dollars for any positive real number \( x \).

There are times when the rules that apply in various cases are closely connected to each other. A classic example is the absolute value function, that is,

\[
f(x) = |x| = \begin{cases} 
  x & \text{if } 0 \leq x, \\
  -x & \text{if } x < 0.
\end{cases}
\]

![Figure 1.2. Graph of \(|x|\).](image)

In this case, \( f(x) = f(-x) \) for all \( x \). When that happens, we say that \( f \) is an even function. For instance, \( g(x) = \cos x \) and \( h(x) = x^2 \) are even functions. There are also functions for which \( -f(x) = f(-x) \) holds for all \( x \). Then we say that \( f \) is an odd function. Examples of odd functions include \( g(x) = \sin x \) and \( h(x) = x^3 \).

There are times when a plain English description of a function is simpler than an algebraic one. For instance, “let \( g \) be the function that sends each integer that is at least 2 into its largest prime divisor” is simpler than describing that function with algebraic symbols (and symbols of formal logic). If the sets \( D \) and \( R \) are not sets of numbers, an algebraic description may not even be possible. An example of this is when \( D \) and \( R \) are both sets of people and \( f(x) \) is the biological father of person \( x \). Note that it is not by accident that we said that \( f(x) \) is the father (and not the son) of \( x \). Indeed, a function must send \( x \) to a unique \( f(x) \). While a person has only one biological father, he or she may have several sons.

Sometimes the rule that sends \( x \) to \( f(x) \) can only be given by listing the value of \( f(x) \) for each \( x \), as opposed to a general rule. For instance, let \( D \) be the set of 200 specific cities in the United States, let \( R \) be the set of all nonnegative real numbers, and for a city \( x \), let \( f(x) \) be the
amount of precipitation that \( x \) had in 2011. Then \( f \) is a function since it sends each \( x \in D \) into an element of \( R \). This function is given by its list of values, not by a rule that would specify how to compute \( f(x) \) if given \( x \).

Finally, functions can also be represented by their graphs. If \( f : D \rightarrow R \) is a function, then let us consider a two-dimensional coordinate system such that the horizontal axis corresponds to elements of \( D \), and the vertical axis corresponds to elements of \( R \). The graph of \( f \) is the set of all points with coordinates \((x, f(x))\) such that \( x \in D \). The requirement that \( f(x) \) is unique for each \( x \) will ensure that no vertical line intersects the graph of \( f \) more than once. This is called the vertical line test.

1.1. Exercises.

(1) For each person \( x \), let \( f(x) \) denote the birthday (day, month, and year) of \( x \). Is \( f \) a function?
(2) For each person \( y \), let \( g(y) \) denote the biological mother of \( y \). Is \( g \) a function? If yes, what is the domain of \( g \) and what is the range of \( g \)?
(3) For two people \( x \) and \( y \), let us say that \( f(x) = y \) if \( y \) is a child of \( x \). Is \( f \) a function?
(4) How many functions are there with domain \( \{A, B, C, D\} \) and range \( \{0, 1\} \)?

For the remaining exercises in this section, all functions are defined on some real numbers.

(5) Let \( f(x) = x + |x| \). Find the domain and the range of \( f \).
(6) Let \( f(x) = (x + 1)/(x - 2) \). Find the domain and the range of \( f \).
(7) Let \( g(x) = x/|x| \). Find the domain and the range of \( f \).
(8) Let \( f(x) = \frac{1}{\sin x} \). Find the domain and the range of \( f \).
(9) Let \( f(x) = \sqrt{\sin^2 x + \cos^2 x} \). Find the domain and the range of \( f \).
(10) Let \( h(x) = \frac{x}{x + 3} + \frac{x + 3}{x} \). Find the domain and the range of \( f \).
(11) Let \( \lfloor x \rfloor \) be the smallest integer \( y \) such that \( x \leq y \). What is the domain and the range of the function \( x \rightarrow \lfloor x \rfloor \)?
(12) Let \( \lfloor x \rfloor \) be defined as in the previous exercise and let \( f(x) = \lfloor x \rfloor - x \). Find the domain and the range of \( x \).
(13) Can the graph of a function intersect a vertical line twice?
(14) Can the graph of a function contain a circle?
(15) Can the graph of a function intersect a horizontal line twice?
(16) Can the graph of a function intersect a horizontal line infinitely many times?
(17) An infinite sequence is an infinite array of numbers \(a_1, a_2, \ldots\). Explain why infinite sequences are, in fact, functions. What is the domain of these functions?
(18) Let \(f(x) = 3x + 2\). Find four points that are on the graph of \(f\). What can be said about the curve determined by those four points?
(19) Let \(f\) and \(g\) be two functions and let us assume that there is exactly one point \((x, y)\) that is on the graph of both \(f\) and \(g\). What is the algebraic meaning of that fact?
(20) Let \(f\), \(g\), and \(h\) be three functions and let us assume that there is no point that is on the graph of all three of them. What is the algebraic meaning of that fact?

2. Classes of Functions

2.1. Power Functions. A power function is a function \(f\) given by the rule \(f(x) = x^a\), where \(a\) is a fixed real number. Note that \(x^{-a} = 1/x^a\), so, for instance, \(x^{-3} = 1/x^3\). The special case of \(a = -1\), that is, the function \(f(x) = 1/x\), is called the reciprocal function. Note that the rule \(g(x) = 1\) for all real numbers \(x\) also defines a power function, one in which \(a = 0\). If \(a = 1/n\), where \(n\) is a positive integer, then the power function \(f\) given by the rule

\[ f(x) = x^a = x^{1/n} = \sqrt[n]{x} \]

is also called a root function.

2.2. Polynomials. A polynomial function is the sum of a finite number of constant multiples of power functions with nonnegative integer exponents, such as the function \(f\) given by the rule \(f(x) = 3x^4 + 2x^2 + 7x - 5\). The domain of these functions is the set of all real numbers. The largest exponent that is present in a polynomial function is called the degree of the polynomial. So the degree of \(f\) in the last example is 4. The real numbers that multiply the power functions in a polynomial are called the coefficients of the polynomial. In the last example, they are 3, 2, 7, and -5.

Some subclasses of polynomial functions have their own names as follows:
• Polynomials of degree 0, such as \( f(x) = 6 \), are called constant functions.
• Polynomials of degree 1, such as \( g(x) = 3x - 2 \), are called linear functions.
• Polynomials of degree 2, such as \( h(x) = x^2 - 4x - 21 \), are called quadratic functions.
• Polynomials of degree 3, such as \( p(x) = x^3 - x^2 + 6x - 2 \), are called cubic functions.

2.3. Rational Functions. A rational function is the ratio of two polynomial functions such as

\[
R(x) = \frac{3x^2 + 4x - 7}{x^3 - 8}.
\]

The domain of a rational function is the set of all real numbers, except for the numbers that make the polynomial in the denominator 0. In the preceding example, the only such real number is \( x = 2 \).

2.4. Trigonometric Functions. Periodicity. The reader has surely encountered the trigonometric functions \( \sin, \cos, \tan, \cot, \sec, \) and \( \csc \) in earlier courses. We will discuss these functions, and their inverses, later in the text. For now, we mention one of their interesting properties, their periodicity. A function \( f \) is called periodic with period \( T > 0 \) if \( f(x) = f(x + T) \) for all \( x \) and \( T \) is the smallest positive real number with this property.

For example, \( \sin \) and \( \cos \) are both periodic with period \( 2\pi \), and \( \tan \) and \( \cot \) are periodic with period \( \pi \). See Figure 1.3 for an illustration. The reader will be asked in Exercise 2.7.1 about the periodicity of \( \sec \) and \( \csc \).

2.5. Algebraic Functions. An algebraic function is a function that contains only addition, subtraction, multiplication, division, and taking roots. For instance, power functions with integer exponents are algebraic functions, since they only use multiplication, though possibly many times. Therefore, polynomials are algebraic functions as well since they are sums of constant multiples of power functions. This implies that rational functions are also algebraic since they are obtained by dividing a polynomial (also an algebraic function) by another one.

The preceding list did not contain all algebraic functions since it did not contain any functions in which roots were involved. So we get additional examples if we include roots, such as the functions given by the rules \( f(x) = \sqrt{x + 3} \), \( g(x) = \sqrt[3]{x} \), \( h(x) = \sqrt{(x + 1)/(x - 1)} \).
2.6. Transcendental Functions. Functions that are not algebraic are called transcendental functions. These include trigonometric functions and their inverses, exponential functions, which are functions that contain a variable in the exponent, such as $f(x) = 2^x$, and their inverses, which are called logarithmic functions. See Figure 1.4 for an illustration. We will discuss these functions in later sections of this chapter. There are many additional examples, which do not have their own names.

2.7. Exercises.

(1) Are secant and cosecant periodic functions? If yes, what is their period?
(2) Can a polynomial be a periodic function?
(3) Are $f(x) = 3x^5 + 7x - 31$ and $g(x) = (2x + 7)/(3x - 1)$ polynomial functions?
(4) Are $f(x) = 2^x$ and $g(x) = \sin^2 x$ power functions?

Figure 1.3. Trigonometric functions.
Figure 1.4. Logarithmic functions.

(5) Are $1/(x + 3)$, $g(x) = (x^2 + 3x + 9)/(x^3 + 1)$, and $h(x) = (\sin x)/(x + 2)$ rational functions?

(6) Show an example of a rational function that is defined for all real numbers.

(7) Show an example of a rational function that is defined for all real numbers except 1, 2, and 3.

(8) Let $f(x) = x^{2/3}$. Is $f$ an algebraic function?

(9) Is $\sin(3x)$ a periodic function? If yes, what is its period?

(10) Is $\cos(1/x)$ a periodic function? If yes, what is its period?

(11) Is $\sin(|x|)$ a periodic function? If yes, what is its period?

(12) Show an example of a periodic function that has period 1.

(13) Let $f(x) = x^{-2/7}$. Is $f$ an algebraic function?

(14) Is $g(x) = (2/3)^x$ an algebraic function?

(15) Show an example of a periodic function with period $\sqrt{\pi}$.

(16) Is $\sin x + \tan x$ a periodic function? If yes, what is its period?

(17) Is $\sin x \tan x$ a periodic function? If yes, what is its period?

(18) Is $\sin^2 x$ a periodic function? If yes, what is its period?

(19) Let $[x]$ be equal to the largest integer $y$ such that $y \leq x$. Now set $f(x) = x - [x]$. Is $f$ a periodic function? If yes, what is its period?

(20) Let $f$ and $g$ be periodic functions defined for all real numbers. Let $f$ have period 5 and let $g$ have period 7. Is $f + g$ a periodic function? If yes, what is its period?
3. Operations on Functions

3.1. Transformations of a Function. We have seen the basic mathematical functions and their graphs in the last section. In this section, we will look at their transformations.

It is easy to see what happens to the graph of a function if we increase or decrease each value of a function by a constant. Indeed, the graph of the function \( g \) given by \( g(x) = f(x) + 5 \) for all \( x \) is simply the graph of the function \( f \) translated by five units to the north. Similarly, the graph of the function \( h \) given by \( h(x) = f(x) - 7 \) is the graph of \( f \) translated by seven units to the south.

Horizontal translations are a little bit trickier. The reader is invited to verify that if \( g \) is the function given by \( g(x) = f(x - 2) \), then the graph of \( g \) is the graph of \( f \) translated by two units to the east, that is, in the positive direction. Indeed, we must substitute a larger number into \( g \) to get the same value as from \( f \). For instance, \( g(8) = f(6) \).

See Figure 1.5 for an illustration. Similarly, if \( h \) is the function given

![Figure 1.5. Horizontal and vertical translations of \( f(x) \).](image-url)
by \( h(x) = f(x + 3) \) for all \( x \), then the graph of \( h \) is the graph of \( f \) translated by three units to the west, that is, in the negative direction.


The effect of multiplication and division on functions can be described similarly. If \( f \) is a function and \( g \) is the function given by \( g(x) = c \cdot f(x) \), where \( c > 1 \) is a real number, then the graph of \( g \) is simply the graph of \( f \) “stretched” vertically by a factor of \( c \). That is, each point on the graph of \( g \) is \( c \) times as far away from the horizontal axis as the corresponding point on the graph of \( f \). It goes without saying that dividing by \( c > 1 \) has the opposite effect. That is, if \( h(x) = f(x)/c \), then the graph of \( h \) is a vertically compressed version of the graph of \( f \). In other words, each point on the graph of \( h \) is \( c \) times as close to the horizontal axis as the corresponding point on the graph of \( f \). See Figure 1.6 for an illustration.

At this point, the reader should stop and think about what happens if \( c < -1 \) is a negative constant. As the reader probably figured out, the stretching or compressing effect will not change (it will only depend on \( |c| \)), but each point on the graph will be reflected through the horizontal axis.

![Figure 1.6](http://www.math.ufl.edu/~mathguy/ufcalc\_book/translations.html)

**Figure 1.6.** Effects of multiplying a function by a constant.

The reader is encouraged to consult the interactive website http://www.math.ufl.edu/~mathguy/ufcalc\_book/squeeze.html for further illustrations.

Horizontal transformations involving multiplication and division are similar to their counterparts involving addition and subtraction in that their effect is the opposite of what one might think at first. If \( c > 1 \)
and $g$ is the function obtained from $f$ by the rule $g(x) = f(cx)$, then the graph of $g$ is the graph of $f$ compressed horizontally by a factor of $c$. That is, each point on the graph of $g$ is $c$ times as close to the vertical line as the corresponding point on the graph of $f$. In other words, if $(x, y)$ is a point on the graph of $f$, then $(x/c, y)$ is a point on the graph of $g$. On the other hand, if $h$ is obtained from $g$ by the rule $h(x) = f(x/c)$, then the graph of $h$ is a horizontally stretched version of the graph of $f$. That is, each point on the graph of $h$ is $c$ times as far from the vertical axis as the corresponding point on the graph of $f$. So if $(x, y)$ is a point on the graph of $f$, then $(cx, y)$ is a point on the graph of $h$. Again, the reader should stop for a minute and think about the graphs of the functions $f(cx)$ and $f(x/c)$ when $c < -1$ is a negative constant.

### 3.2. Combining Two Functions.

If $f$ and $g$ are two functions, then their sum, difference, and product are defined wherever both $f$ and $g$ are defined. That is, the domain of $f + g$, $f - g$, and $fg$ is the intersection of the domains of $f$ and $g$. Furthermore, $(f + g)(x) = f(x) + g(x)$, $(f - g)(x) = f(x) - g(x)$, and $(fg)(x) = f(x)g(x)$. Figure 1.7 illustrates the sum of two functions. We have to be just a little bit more careful with $f/g$, since this function is not defined when $g(x) = 0$, even if $x$ is in the domain of both $f$ and $g$. So the domain of $f/g$ is the intersection of the domain of $f$ and the domain of $g$, with the exception of the points $x$ satisfying $g(x) = 0$. For each point of this domain, $(f/g)(x) = f(x)/g(x)$.

![Figure 1.7. Adding two functions together.](image)

If the range of $f$ is part of the domain of $g$, then we can compose $f$ and $g$ by first applying $f$ and then $g$. The function we obtain in this
way sends \( x \) to \( g(f(x)) \) and is called the composition of \( f \) and \( g \). It is denoted by \( f \circ g \). Note that in \( f \circ g \), first \( f \), and then \( g \) is applied.

**Example 1.1.** Let \( \mathbb{R} \) be the set of all real numbers. If \( f \) and \( g \) are both functions from \( \mathbb{R} \) to \( \mathbb{R} \) and \( f(x) = x^2 \) and \( g(x) = x + 1 \), then

\[
(f \circ g)(x) = g(f(x)) = x^2 + 1,
\]

while

\[
(g \circ f)(x) = f(g(x)) = (x + 1)^2 = x^2 + 2x + 1.
\]

Note that \( f \circ g \) and \( g \circ f \) are, in general, different functions.

### 3.3. Exercises.

1. Sketch the graph of \( f(x) = x^2 \), \( g(x) = (x - 3)^2 \), and \( h(x) = (2x + 5)^2 \).
2. Sketch the graph of \( f(x) = (x + 4)^2 \) and \( g(x) = x^2 + 4 \).
3. Sketch the graph of \( f(x) = |x + 5| \) and \( g(x) = |x| + 5 \).
4. Sketch the graph of \( f(x) = |x| + 1 \) and \( g(x) = |x + 1| \).
5. Sketch the graph of \( f(x) = \sin(x/2) \) and \( g(x) = (\sin x)/2 \).
6. Sketch the graph of \( f(x) = |\sin x| \).
7. Sketch the graph of \( f(x) = \sqrt{x} \) and \( g(x) = 1/\sqrt{x} \).
8. Sketch the graph of \( f(x) = e^{x-3} \) and \( g(x) = e^x - 3 \).
9. Sketch the graph of \( f(x) = \ln(x^2) \) and \( g(x) = (\ln x)^2 \).
10. Sketch the graph of \( f(x) = \cos(x + \pi) \) and \( g(x) = \pi + \cos x \).
11. Sketch the graph of \( f(x) = \sqrt{x + 10} \) and \( g(x) = \sqrt{x} + 10 \).
12. Sketch the graph of \( f(x) = \cos(2x) \), \( g(x) = \sin(x - 2) \), and \( h(x) = 3\tan x \).
13. Sketch the graph of \( f(x) = 2\cos x \), \( g(x) = 2(\sin x) - 2 \), and \( h(x) = \tan(3x) \).
14. Show examples for \( f \) and \( g \) when \( g \circ f \) is defined for all real numbers, but \( f \circ g \) is not.
15. Show examples when \( f \circ g = g \circ f \).
16. Show examples when \( (f \circ g) \circ h = f \circ (g \circ h) \).
17. Show examples when \( (f \circ g) \circ h \neq f \circ (g \circ h) \).
18. Sketch the graph of \( g(x) = \sin(x - \pi/4) \).
19. Let \( f(x) = \sin x \) and \( g(x) = x^2 \). Determine \( f \circ g \) and \( g \circ f \) and sketch their graph.
20. Let \( f(x) = \cos x \), let \( g(x) = e^x \), and let \( h(x) = \ln x \). Determine \( f \circ (g \circ h) \) and \( (f \circ g) \circ h \).
4. Viewing the Graphs of Functions

4.1. The Graph of Function. The graph of a function \( f \) is the set

\[
\{(x, f(x)) | x \in D(f)\}.
\]

The graph of a function is a good way of visually describing what a function does. Today, we have plenty of advanced tools, such as computer software packages and graphing calculators, to study the graph of functions. In this section, we point out a few of the common mistakes in using these tools.

In order to facilitate the discussion, let us agree on some terminology. If the domain of \( f \) contains an interval \( I \) and for all real numbers \( x \) and \( x' \) in \( I \), it is true that \( x < x' \) implies \( f(x) < f(x') \), then we say that \( f \) is increasing on \( I \). Visually, this means that the graph of \( f \) goes roughly from the southwest to the northeast while \( x \in I \). Similarly, if, for all real numbers \( x \) and \( x' \) in \( I \), it is true that \( x < x' \) implies \( f(x) > f(x') \), then we say that \( f \) is decreasing on \( I \). In terms of the graph of \( f \), this means that the graph goes roughly from the northwest to the southeast.

If we simply ask a computer or graphing calculator to plot the graph of a function without specifying the interval \([x_1, x_2]\) in which the value of \( x \) can range, we may get an error message, or the computer may simply substitute default values for \( x_1 \) and \( x_2 \). For example, the software package Maple 13 uses the default values \( x_1 = -10 \) and \( x_2 = 10 \). The interval \([x_1, x_2]\) is often called the viewing window. See Figure 1.8 for an illustration.

![Figure 1.8](image)

**Figure 1.8.** Viewing \( g(x) = 4x^3 + 9x^2 + 6x + 1 \) with viewing window \([-10,10]\) and \([-1,0]\).

We have to be careful, however, since not all viewing windows are appropriate for all functions, and choosing an inappropriate viewing window may cause misleading results.
For functions like \( f(x) = x \), \( g(x) = |x| \), or \( h(x) = x^2 + 3 \), the viewing window \([-10, 10]\) is appropriate as the behavior of these functions outside that window is similar to their behavior inside the window.

Now let \( f(x) = (x + 10)^2 \). In this case, using the viewing window \([-10, 10]\), we get the graph of an increasing function. That is misleading since \( f \) is decreasing on the interval \((-\infty, -10]\). So, in this case, a viewing window that starts at a point \( x_1 < -10 \) is necessary.

This problem becomes more difficult if we are dealing with functions that change from increasing to decreasing many times, perhaps in an irregular fashion and perhaps far away from the origin. For this reason, it is worth noting that if \( f \) is a polynomial function of degree \( n \), then it cannot change directions more than \( n - 1 \) times. If we found all \( n - 1 \) direction changes, then we can be sure that we did not miss any of them. We will return to this topic in a later chapter, when we discuss the derivative of a function.

The preceding example showed why selecting a viewing window that is too small can be misleading. The next example shows why a viewing window that is too large can also mislead us. Plot the graph of the function \( g(x) = 4x^3 + 9x^2 + 6x + 1 \). Using the default viewing window \([-10, 10]\), or some window containing that one, many software packages will show a graph that increases everywhere and disappears in a small interval to the left of 0. This should raise our suspicion that the program does not properly display the graph of \( g \) around 0. Indeed, \( g \) is defined for all real numbers, so its graph should not disappear anywhere. Taking a closer look, that is, changing the viewing window to \([-1, 1]\), we see a function that is actually decreasing between \( x = -1 \) and \( x = -1/2 \).

Trigonometric functions, with their periodicity, are particularly good examples to demonstrate what software packages can and cannot do. The reader is encouraged to plot the graph of the functions \( \sin x \), \( \cos(2x) \), \( \tan(x/4) \), and, finally, \( \sin(1/x) \) and explain the obtained graphs. In particular, the reader should try to explain why, for \( \sin(1/x) \), the choice of the viewing window is not important as long as it contains \( x = 0 \).

Applications of graphical representations of functions include counting the solutions of certain equations even when we cannot explicitly solve those equations, and finding asymptotes. (Note that a solution obtained by simply viewing the graphical representation of a function is not mathematically rigorous, but it can provide a first step for a more rigorous solution.) A horizontal asymptote of a function \( f \) is a horizontal line \( y = a \) so that the values of \( f(x) \) are never equal to \( a \),
but get closer and closer to it as \( x \) gets closer and closer to positive infinity or negative infinity. A vertical asymptote of \( f \) is a line \( x = b \) so that the function \( f \) is not defined at \( x = b \), but as \( x \) gets closer and closer to \( b \), the values of \( f(x) \) get closer and closer to infinity, or negative infinity. For instance, the function \( f(x) = 1/x \) has a horizontal asymptote at \( y = 0 \), and a vertical asymptote at \( x = 0 \). We will make these notions more precise in the next chapter, when we introduce the concept of limits. For now, we can use a graphing software package to find asymptotes, as you will be asked to do in the exercises.

4.2. Exercises. In the following exercises, use a graphing software package with the appropriate viewing window to find the number of solutions (among real numbers) for the given equation. Also find the intervals on which the left-hand side is increasing and on which the left-hand side is decreasing. Approximate the endpoints of these intervals to one decimal.

(1) \( x^4 - x + 1 = 0 \).
(2) \( x^4 - 1 = 0 \).
(3) \( x^3 - 6x + 1 = 0 \).
(4) \( x^3 + x^2 - 1 = 0 \).
(5) \( x^4 - 4x^2 + 1 \).

In the following exercises, use a graphing software package with the appropriate viewing window to find the number of solutions (among real numbers) for the given equation.

(6) \( x^3 - x^2 - 1 = 0 \).
(7) \( x^2 - x - 7 = x^3 - 1 \).
(8) \( x = \sin x \).
(9) \( x^2 = \sin x \).
(10) \( x/2 = \cos x \).
(11) \( x + 2 = 2^x \).

In the following exercises, use a graphing software package to decide if the given function has a vertical or horizontal asymptote.

(12) \( f(x) = (x + 3)/(x + 2) \).
(13) \( f(x) = (x - 7)/(x + 5) \).
(14) \( g(x) = 1/(2 - x) \).
(15) \( h(x) = x + (1/x) \).
(16) \( h(x) = x^2 - (1/x^2) \).
(17) \( s(x) = \sqrt{(x - 4)/(x - 3)} \).
(18) \( s(x) = \sqrt{(x + 5)/(x - 7)} \).
(19) \( z(x) = (x^2 + 1)/(2x^2 - 3) \).
(20) \( z(x) = (x^2 + 4x + 5)/(x + 3) \).
5. Inverse Functions

The inverse \( f^{-1} \) of a function \( f : A \to B \) “undoes” what \( f \) did. That is, if \( f(x) = y \), then \( f^{-1}(y) = x \), so \( f \) sends \( x \) to \( y \), while \( f^{-1} \) sends \( y \) back to \( x \). It goes without saying that this \( f^{-1} \) will only be a function if \( f^{-1}(y) \) is unambiguous, that is, when there is only one \( x \in A \) so that \( f(x) = y \). In that case, and only in that case, it is clear that \( f^{-1}(y) = x \).

Let us now formalize these concepts.

**Definition 1.1.** A function \( f : A \to B \) is called **one-to-one** if it sends different elements into different elements, that is, if \( x \neq x' \) implies that \( f(x) \neq f(x') \).

One-to-one functions are also called **injective functions** or **injections**. Visually, no horizontal line can intersect the graph of a one-to-one function more than once.

For instance, if \( A \) and \( B \) are both the set of real numbers, then \( f(x) = x \) and \( g(x) = x^3 \) are both one-to-one, but \( h(x) = x^2 \) is not.

**Definition 1.2.** Let \( f \) be a one-to-one function with domain \( A \) and range \( B \). Then the **inverse** of \( f \) is the function \( f^{-1} : B \to A \) given by \( f^{-1}(y) = x \) if \( f(x) = y \).

**Example 1.2.** Let \( A \) and \( B \) both be the set of all real numbers. Let \( f : A \to B \) be given by \( f(x) = 2x + 7 \). Then \( f^{-1}(y) = (y - 7)/2 \).

**Solution:** If \( f(x) = y \), then \( y = 2x + 7 \), so \( y - 7 = 2x \) and so \( (y - 7)/2 = x \). As \( x = f^{-1}(y) \), it follows that \( f^{-1}(y) = (y - 7)/2 \).

The preceding example shows a general strategy for finding the inverse of a function. Write the equation \( f(x) = y \), with the appropriate algebraic expression replacing \( f(x) \). Then solve for \( x \). If there is more than one solution, then \( f \) is not one-to-one, and so it has no inverse function. If there is one solution, then that expression is the value of \( f^{-1}(y) \).

**Example 1.3.** If \( A \) is the set of positive real numbers, \( B \) is the set of real numbers that are larger than 1, and \( f : A \to B \) is given by \( f(x) = x^2 + 1 \), then \( f^{-1}(y) = \sqrt{y - 1} \).

**Solution:** We have \( f(x) = x^2 + 1 = y \). So \( x^2 = y - 1 \), and because we know that \( x \) is positive and \( y > 1 \), we can take the square root of both sides, leading to \( x = \sqrt{y - 1} \). Hence, \( f^{-1}(y) = \sqrt{y - 1} \).

Note that the graphs of \( f \) and \( f^{-1} \) are reflected images of each other through the line \( y = x \) as illustrated in Figure 1.9.
Finally, we point out that if $f$ is a one-to-one function with domain $A$ and range $B$, then $f \circ f^{-1}$ is the identity function of $A$ and $f^{-1} \circ f$ is the identity function of $B$.

![Figure 1.9](image.png)

**Figure 1.9.** $f(x)$ and $f^{-1}(x)$ are symmetric about the identity function $x$.

For instance, using the functions of Example 1.3, for all positive real numbers $x$, the identity $(f \circ f^{-1})(x) = \sqrt{x^2 + 1} - 1 = \sqrt{x^2} = x$ holds, and for all $y > 1$, the identity $(f^{-1} \circ f)(y) = (\sqrt{y - 1})^2 + 1 = y - 1 + 1 = y$ holds.

### 5.1. Logarithmic Functions.

If a function contains only additions, subtractions, multiplications, and divisions, then its inverse is often easy to compute. Power functions, that is, functions of the form $f(x) = x^\alpha$, where $\alpha$ is a real number, are not much more difficult. However, what is the inverse of an exponential function?

Let $f(x) = 2^x$. It is easy to see, by plotting the graph of $f$ or otherwise, that $f$ is a one-to-one function whose domain is the set of all real numbers and whose range is the set of all positive real numbers. So the inverse of $f$ is a function from the set of positive reals to the set of all reals. But what is that inverse function $f^{-1}$? By the definition of inverse functions in general, this is the function that sends $2^x$ to $x$ for all positive real numbers $2^x$. In particular, $f^{-1}(2) = 1$, $f^{-1}(4) = 2$, $f^{-1}(32) = 5$, and $f^{-1}(1/2) = -1$. That is, $f^{-1}(y)$ tells us to what power we have to raise 2 if the result is to be $y$. This important concept has its own name.

**Definition 1.3.** Let $m$ be a positive real number. Then the inverse of the function $f(x) = m^x$ is called the logarithmic function with base $m$, and is denoted by $\log_m$. 
So if \( f(x) = x^m = y \), then \( \log_m(y) = x \). For instance, \( \log_2(64) = 6 \), \( \log_3(81) = 4 \), \( \log_5(1/25) = -2 \), and \( \log_{0.5}(16) = -4 \).

Logarithmic functions satisfy certain rules that are very similar to those satisfied by exponential functions and can, in fact, be deduced from them. These are

(I) \( \log(xy) = \log x + \log y \).

(II) \( \log(x/y) = \log x - \log y \).

(III) \( \log(x^a) = a \log x \).

(IV) \( \log \sqrt{x} = \frac{\log x}{2} \).

(V) \( a^{\log_a x} = x \).

(VI) \( \log_a(a^x) = x \).

The last two rules simply express the fact that the functions \( f(x) = a^x \) and \( f^{-1}(y) = \log_a(y) \) are inverses of each other, so their composition is an identity function.

If we know the logarithm of a number in a base and want to compute it in another base, we can do so using the following theorem.

**Theorem 1.1.** For positive real numbers \( a, b, \) and \( x \), we have

\[
\log_a x = \frac{\log_b x}{\log_b a}.
\]

**Proof.** Start with the identity

\( x = a^{\log_a x} \).

Now take the logarithm of base \( b \) of both sides to get

\( \log_b x = \log_a x \log_b a \).

Now divide both sides by \( \log_b a \) to get the identity of the theorem. \( \Box \)

**Example 1.4.** We can use Theorem 1.1 to compute \( \log_{16}(256) \) from \( \log_2(256) \) as follows:

\[
\log_{16}(256) = \frac{\log_2(256)}{\log_2(16)} = \frac{8}{4} = 2.
\]

So if a calculator or computer can provide the logarithm of all positive real numbers in **one** base, it can compute the logarithm of any positive real number in any base. For this reason, many calculators and computers are programmed to work primarily with logarithms of one given base, namely of base \( e \), where \( e \simeq 2.718 \) is an irrational number that will be formally defined in Chapter 2.

The logarithm of base \( e \) is so important that it has its own name, **natural logarithm**, and its own notation, \( \ln \). So \( \ln x = \log_e x \).
5.2. Inverses of Trigonometric Functions. Basic trigonometric functions, such as \( \sin, \cos, \) and \( \tan \), are very important in calculus, so it is no surprise that their inverse functions are important as well. However, we have to be precise when we define them since trigonometric functions are not one-to-one. In fact, they are periodical, of period \( 2\pi \) or \( \pi \), and so they take every value in their range infinitely often.

In order to get around this difficulty, we will restrict our trigonometric functions to just a short interval, in which they are one-to-one, and define their inverses based on that restriction.

For instance, consider \( \sin \) as a function whose domain is \([ -\pi/2, \pi/2 ]\). In that interval, \( \sin \) is a one-to-one function (since it is increasing), and its range is the interval \([-1, 1]\). See Figure 1.10 for an illustration. So the inverse of \( \sin : [-\pi/2, \pi/2] \to [-1, 1] \) is the function \( \sin^{-1} : [-1, 1] \to [-\pi/2, \pi/2] \). That is, if \( y \in [-1, 1] \), then \( \sin^{-1} y \) is the (only) \( x \in [-\pi/2, \pi/2] \) for which \( \sin x = y \). For instance, \( \sin^{-1}(1/2) = \pi/6 \), while \( \sin^{-1}(0) = 0 \) and \( \sin^{-1}(\sqrt{2}/2) = \pi/4 \). Figure 1.11 shows the graph of \( \sin^{-1} x \).

The inverses of the other trigonometric functions are defined similarly, just the intervals to which we restrict the functions (in order to make them one-to-one) can change.

That is, \( \cos^{-1} \) is the inverse function of the \( \cos \) function that is restricted to the interval \([0, \pi]\). So \( \cos^{-1} \) is a function with domain \([-1, 1]\) and range \([0, \pi]\). Similarly, \( \tan^{-1} \) is the inverse function of the \( \tan \) function that is restricted to the interval \((-\pi/2, \pi/2)\). Its domain is the set of all real numbers, and its range is the interval \((-\pi/2, \pi/2)\). See Figure 1.12 for illustrations.

The inverse functions of \( \cot, \sec, \) and \( \csc \), while not used often, can also be defined analogously.

5.3. Exercises.

(1) Is there a function \( f \) defined on all positive real numbers for which \( f^{-1} = f \)?
Figure 1.11. Graph of $\sin^{-1} x$.

Figure 1.12. Graphs of $\cos x$ and $\tan x$ with their inverses.
(2) If we are given \( \log_a x \), how can we compute \( \log_{1/a} x \)?

(3) For which values of \( a \) is \( \log_a \) an increasing function, and for which values of \( a \) is it a decreasing function?

(4) What is the geometric connection between the graphs of \( f \) and \( f^{-1} \)?

(5) What is the value of \( \tan^{-1} 1 \)?

(6) Find all real numbers \( y \) such that \( \tan^{-1} y = \cot^{-1} y \).

(7) Is it true that if \( g \) is the inverse function of the one-to-one function \( f \), then \( g \) is one-to-one?

(8) Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be defined by \( f(x) = |x| \). Is \( f \) a one-to-one function?

(9) Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be defined by \( f(x) = x^5 \). Is \( f \) a one-to-one function?

(10) Let \( f : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) be defined by \( f(x) = x^2 \). Is \( f \) a one-to-one function?

(11) Is \( f(x) = \log_a x \) a one-to-one function on the set of all positive real numbers?

(12) Express \( x \) in terms of \( y \) if \( \log_a (\log_a x) = y \).

(13) Let us assume that \( f : \mathbb{R} \rightarrow \mathbb{R} \) is a strictly increasing function, that is, if \( x < y \), then \( f(x) < f(y) \). Can we conclude that \( f : \mathbb{R} \rightarrow \mathbb{R} \) is a one-to-one function?

(14) Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be a function that has an inverse function \( f^{-1} : \mathbb{R} \rightarrow \mathbb{R} \). If \( f \) is strictly increasing, can we conclude that \( f^{-1} \) is strictly increasing?

(15) The hyperbolic sine function is defined by

\[
\sinh x = (e^x - e^{-x})/2.
\]

Prove that \( \sinh^{-1} y = \ln(y + \sqrt{y^2 + 1}) \) for all real numbers \( y \).

(16) The hyperbolic cosine function is defined by

\[
\cosh x = (e^x + e^{-x})/2.
\]

Prove that for all real numbers \( x \), the inequality \( \cosh x \geq 1 \) holds.

(17) Prove that \( \cosh^{-1} y = \ln(y + \sqrt{y^2 - 1}) \) for all real numbers \( y \geq 1 \).

(18) Is \( \cosh : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) a one-to-one function?

(19) The hyperbolic tangent function is defined by

\[
tanh x = \sinh x / \cosh x.
\]

Prove that for all real numbers \( x \), the inequality \( |\tanh x| < 1 \) holds.

(20) Prove that if \( |y| < 1 \), then \( \tanh^{-1} y = \frac{1}{2} \ln \left( \frac{1+y}{1-y} \right) \).
6. The Velocity Problem and the Tangent Problem

6.1. The Velocity Problem. Let us assume that a car was on the road from 3:00 p.m. to 5:00 p.m. on a given afternoon, and it traveled a distance of 100 miles, all due west. From the data, it is easy to compute the average speed of the car by the formula

$$v = \frac{s}{t},$$

where \( t \) is the time passed, \( s \) is the distance covered in time \( t \), and \( v \) is the average speed for the given time period. In physics, when the direction in which an object is moving is taken into account, we talk about velocity instead of speed, hence the abbreviation \( v \). In the given example, all travel was in one direction (west), so there is no danger of confusion, and we can use either word. Let us assume that time is measured in hours and distance is measure in miles.

Then Equation (1.1) yields

$$v = \frac{100 \text{ mi}}{2 \text{ hr}} = 50 \frac{\text{mi}}{\text{hr}},$$

so the average velocity of the car for the given 2-hour period is 50 miles per hour.

The car probably did not cover the entire distance at its average velocity. For various traffic-related or other reasons, it sometimes may have gone faster or slower. If we want to know its average velocity for the time period between 4:00 p.m. and 4:10 p.m., then we need to know the distance it covered in that time period. If that distance is 10 miles, then we conclude that in that 10-minute time period, the average velocity of the car was

$$v = \frac{10 \text{ mi}}{1/6 \text{ hr}} = 60 \frac{\text{mi}}{\text{hr}}.$$ 

If we want more precise information, like the average velocity of the car between 4:02 p.m. and 4:05 p.m., we can proceed similarly, decreasing the value of both the numerator and the denominator of the fraction \( s/t \). However, what if we want to know the instantaneous velocity of the car in a given moment, such as exactly at 4:02:23 p.m. (and not in the second that passed between 4:02:23 p.m. and 4:02:24 p.m.)? In that case, a direct application of Equation (1.1) is impossible, because the denominator \( t \) is equal to 0. The numerator \( s \) is also equal to 0, since the car needs time to cover any distance; if it is given no time, it will cover no distance.

In this section, we will not give a completely formal answer to the problem of defining instantaneous velocity; we will leave that task to
an upcoming section. However, we will say the following. The instantaneous velocity of a car in a given moment \( m \) can be approximated by choosing smaller and smaller time periods containing \( m \) and computing the average speed of the car for those time periods. These averages will approximate the instantaneous velocity.

6.2. The Tangent Problem. The problem of finding the instantaneous velocity of a moving object is simply a special case of a much more general problem, that of finding the slope of a tangent line to a curve at a given point.

In the previous problem, the distance the car covered can be viewed as a function of the time that passed since the car started moving. So \( s(t) \) is the distance covered from the moment when the car started moving to the moment \( t \) hours later. In order to compute the average velocity for the time period from \( t_1 \) to \( t_2 \), we simply compute the value of the fraction

\[
\frac{s(t_2) - s(t_1)}{t_2 - t_1}.
\]

This fraction is precisely the slope of the line that intersects the graph of the function \( s \) at points \((t_1, s(t_1))\) and \((t_2, s(t_2))\). If we choose \( t_1 \) and \( t_2 \) closer and closer together, then these points will get closer and closer together as well. Finally, if we set \( t_1 = t_2 \), then we will not immediately know the slope of the line that touches the graph of \( s \) at the point \((t_1, s(t_1))\) since we will know only one, not two, point of this line. However, and this will be made more precise in the next section, the slope we are looking for will be approximated by the sequence of slopes of the lines that we got when we chose \( t_1 \) and \( t_2 \) closer and closer together.

Finally, we point out that there is nothing magical about the function \( s(t) \) here. We could consider any function \( f : \mathbb{R} \to \mathbb{R} \), and ask what the slope of the tangent line to this curve is at the point \((x, f(x))\).

6.3. Exercises.

1. A car travels 1 hour at a speed of 60 miles per hour, then 2 hours at a speed of 45 miles per hour. What is the average speed of the car during this 3-hour period?
2. Consider the car of the previous exercise. What is its average speed during the first 2 hours of its trip?
3. I drove at 40 miles per hour for 2 hours. How fast do I have to drive in my third hour if I want to reach an average speed of 45 miles per hour for my 3-hour drive?
4. A car travels 300 miles on a given day. During the first 100 miles, the car travels at a speed of 40 miles per hour, during the second 100 miles, it travels at a speed of 50 miles per hour, and during the third 100 miles, it travels at a speed of 60 miles per hour. What is the average speed of the car for the entire 300-mile trip?

5. Tim has ridden his bicycle to school, covering a 5-mile distance in half an hour. Can we conclude that there was a segment of his ride for which his average speed was more than 10 miles per hour?

6. Jim has driven his car for 3 days in different conditions. On the first day, he was able to drive 20 miles per gallon of fuel used, for a total of 200 miles. On the second day, he drove 25 miles per gallon of fuel used, for a total of 275 miles. On the third day, he drove 24 miles per gallon of fuel used, for a total of 240 miles. What was his average number of miles driven per gallon of fuel for the entire 3-day trip?

7. A ball is thrown vertically in the air. In \( t \) seconds, its height (in meters) is given by the function \( h(t) = 50t - 20t^2 \). What is the average velocity of the ball during its first 2 seconds of motion?

8. Consider the ball of the previous exercise. What is the average velocity of the ball between \( t = 1 \) and \( t = 2 \)?

9. Consider the ball of the previous exercise. What is the average velocity of the ball between \( t = 1.4 \) and \( t = 1.6 \)?

10. Give a reasonable estimate of the velocity of the ball of the previous exercise in the moment \( t = 1.5 \).

11. Consider the function \( f(x) = x^2 \). Can you find two points \( P \) and \( Q \) on the graph of \( f \) such that the slope of the line \( PQ \) is between 0 and 0.01?

12. Let \( f(x) = \sqrt{x} \) and let \( P = (1,1) \). Find the slope of the three lines that connect \( P \) to the points \((4,2), (2.25,1.5)\), and \((1.44,1.2)\).

13. Let \( f \) be as in the previous exercise. Find the slope of the three lines connecting \( P = (1,1) \) to the points \((0.25,0.5), (0.64,0.8)\), and \((0.81,0.9)\).

14. Consider the results of the two preceding exercises. Do you see a trend?

15. Let \( g(x) = e^x \) and let \( P = (0,1) \). Find the slope of the three lines connecting \( P \) to the points \((-1,e^{-1}), (1,e)\), and \((\ln 2,2)\).
(16) Consider the function \( f(x) = x^2 \). Let \( P = (1, 1) \). Can you find a point \( Q \) on the graph of \( g \) such that the slope of the line \( PQ \) is 2?

(17) Consider the function \( g(x) = x^3 \). Let \( P = (1, 1) \). Can you find a point \( Q \) on the graph of \( g \) such that the slope of the line \( PQ \) is between 1 and 1.01?

(18) Consider the function \( f(x) = \frac{1}{x} \). Choose two points \( P \) and \( Q \) of the graph of \( f \) such that \( P \neq Q \) and the \( x \) coordinates of \( P \) and \( Q \) are small and positive. What can be said about the slope of the line \( PQ \)?

(19) Consider the function \( f(x) = e^x \). Can you find two points on the graph of \( f \) such that the slope of the line \( PQ \) is negative? Explain your answer.

(20) Consider function \( f(x) = 1 - \ln x \). Can you find two points on the graph of \( f \) such that the slope of the line \( PQ \) is positive? Explain your answer.
7. The Limit of a Function

7.1. Two-Sided Limits. Consider the function given by the rule \( f(x) = \frac{1}{1 + x} \). Let us compute the values of \( f(x) \) for various real numbers \( x \) that are close to 0. We find that

- \( f(1) = \frac{1}{2} \),
- \( f(1/2) = \frac{2}{3} \),
- \( f(1/3) = \frac{3}{4} \), and, in general,
- \( f(1/n) = n/(n + 1) \).

Similarly, for negative values of \( x \), we get

- \( f(-1/2) = 2 \),
- \( f(-1/3) = 3/2 \),
- \( f(-1/4) = 4/3 \), and, in general,
- \( f(-1/n) = n/(n - 1) \).

What we see is that if \( x \) gets close to 0 (from either side), then \( f(x) \) gets close to \( f(0) = 1 \). In fact, we can get \( f(x) \) to be as close to \( f(0) = 1 \) as we want; all we need to do is to choose \( x \) sufficiently close to 0. Indeed, looking at the previous examples, we conclude that if \( 0 < x < 1/n \), then \( n/(n + 1) < f(x) < 1 \), and if \( -1/n < x < 0 \), then \( 1 < f(x) < n/(n - 1) \). So for instance, if we want \( f(x) \) to be closer than \( \frac{1}{1000} \) to 1, then any choice of \( x \) in the interval \( [0, \frac{1}{999}] \) or any choice of \( x \) in the interval \( (-\frac{1}{1001}, 0] \) will work. That is, any choice of \( x \) in the interval \( (-\frac{1}{1001}, \frac{1}{999}) \) will imply that \( |f(x) - f(0)| < 0.001 \).

This phenomenon, that is, the fact that there exists an interval such that, for each real number in that interval, the value of \( f(x) \) is closer to \( f(0) \) than a prescribed bound is so important in mathematics that it has its own name.

**Definition 2.1.** Let \( f : \mathbb{R} \to \mathbb{R} \) be a function and let \( a \) be a real number. We say that the limit of \( f \) in \( a \) is the real number \( L \) if the values of \( f(x) \) get arbitrarily close to \( L \) and stay arbitrarily close to \( L \) when \( x \) is suitably close to \( a \) without being equal to \( a \).

The fact that the limit of \( f \) in \( a \) is \( L \) is expressed by the notation

\[
\lim_{x \to a} f(x) = L.
\]
So, if \( f \) is the starting example of this section, then \( \lim_{x \to 0} f(x) = 1 \).

Note that the definition of \( \lim_{x \to a} f(x) \) requires that \( f(x) \) stay close to \( L \) when \( x \) is close to \( a \), regardless of which of \( x \) or \( a \) is larger. That is, \( f(x) \) has to be close to \( L \) if \( x \) is a little bit less than \( a \), and \( f(x) \) has to be close to \( L \) if \( x \) is a little bit more than \( a \), though \( f(x) \) does not have to be close to \( L \) if \( x = a \).

Several comments are in order. First, \( \lim_{x \to a} g(x) \) does not always exist.

Example 2.1. Let

\[
g(x) = \begin{cases} 
1 & \text{if } 0 \leq x, \\
0 & \text{if } x < 0.
\end{cases}
\]

Then the limit of \( g \) at \( a = 0 \) does not exist. Indeed, no matter how small an interval \( I \) we take around the point \( a = 0 \), that interval \( I \) will contain some positive and some negative real numbers. Hence, the values of \( g(x) \) will sometimes equal 1 and sometimes equal 0 for \( x \in I \), no matter how small \( I \) is. There is no number \( L \) such that 0 and 1 are arbitrarily close to it—in fact there is no number such that both 0 and 1 are both closer than 0.5 to it. So \( \lim_{x \to 0} g(x) \) does not exist.

Second, if \( \lim_{x \to 0} f(x) \) exists, it is unique; that is, \( f \) cannot have two different limits at any given point \( a \). Let us illustrate this using the introductory example of this section, the function \( f(x) = \frac{1}{1+x} \). We have seen that \( \lim_{x \to 0} f(x) = 1 \). Indeed, we saw that the values of \( f(x) \) can get arbitrarily close to 1 if the real numbers \( x \) are chosen from a suitably small interval around 0. At this point, one could ask the following question. If 1 satisfies the requirements to be the \( \lim_{x \to 0} f(x) \), why does 1.0001 not? After all, what is close to 1 is also close to 1.0001.

In order to answer this question, we must have a good understanding of the definition of limits. That definition says that if \( \lim_{x \to 0} f(x) = L \), then the values of \( f(x) \) will get arbitrarily close to \( f(0) \) if \( x \) is chosen from a suitably small interval around 0. The key word in the previous sentence is arbitrarily. While 1.0001 is close to 1, it is not arbitrarily close to 1; it is exactly 0.0001 away. And that is a problem, since we have seen at the beginning of this chapter that, as \( x \) approaches 0, the values of \( f(x) \) will get arbitrarily close to 1. In particular, if \( x \) is close enough to 0, then \( f(x) \) will be closer than \( \frac{1}{10^6} \) to 1, but then it cannot also be closer than \( \frac{1}{10^6} \) to 1.0001.

An analogous argument shows that no function can have two different limits at any one point.

Sometimes it can happen that \( h \) is not even defined in \( a \), but \( \lim_{x \to 0} h(x) \) still exists. Note that the fact that \( h(a) \) is not defined
is not a problem since the definition of limits specifically states that $x$ should not be equal to $a$ anyway.

**Example 2.2.** Let $h(x) = (x^2 - 9)/(x - 3)$. Then $h$ is defined for all real numbers except $x = 3$. Still, $\lim_{x \to 3} h(x) = 6$. In particular, $\lim_{x \to 3} h(x)$ exists.

See Figure 2.1 for an illustration.

**Solution:** If $x \neq 3$, then

$$f(x) = \frac{x^2 - 9}{x - 3} = \frac{(x + 3)(x - 3)}{x - 3} = x + 3.$$ 

So if we want $f(x) = x + 3$ to be closer to 6 than a given distance $a$, then all we have to do is to choose $x$ such that $|x - 3| < a$.  

At this point, the reader should test his or her understanding of the material by finding $\lim_{x \to -2}((x^2 + 3x + 2)/(x + 2))$.

Sometimes, limits are not easy to determine. Plotting the graph of the function $h(x) = (\sin x)/x$, we are led to believe that

$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$ 

See Figure 2.2 for an illustration.
However, we have not yet learned the techniques to rigorously prove this. Plotting the graph of the function or producing more numerical data should not be considered as a complete answer, since, as \( x \) approaches 0, eventually \( x \) and \( \sin x \) will get so small that the computer will no longer manipulate them, or their ratio, accurately.

Finally, we point out that in the definition of the limit, the requirement that \( f(x) \) get close to \( L \) and stay close to \( L \) is important. Consider the function \( f(x) = \sin(1/x) \) around \( x = 0 \). As \( x \) approaches 0, the value of \( 1/x \) will increase very fast, and so it will equal a multiple of \( \pi \) many times. All those times, \( f(0) = 0 \) will hold, so \( f(x) \) will be as close to 0 as possible. However, \( \lim_{x \to 0} f(x) \) does not exist, since \( f(x) \) will take all other values in the interval \([-1, 1]\) infinitely often as well as \( x \) approaches 0. So the value of \( f(x) \) will not stay arbitrarily close to 0, no matter how close \( x \) is to 0. See Figure 2.3 for an illustration.
7.2. The Precise Definition of Limits. It is time for us to give a precise mathematical definition of limits. The advantage of this formal definition is that we can finally do away with the words *arbitrarily close* and *sufficiently close*. The price to pay for that is that we have to use more notation.

**Definition 2.2.** Let $f$ be a function defined on some open interval that contains the real number $a$, with the possible exception of $a$ itself. Then we say that the limit of $f$ at $a$ is $L$, denoted by \( \lim_{x \to a} f(x) = L \), if, for all $\epsilon > 0$, there exists $\delta > 0$ such that if $|x - a| < \delta$, then $|f(x) - L| < \epsilon$.

See Figure 2.4 for an illustration.

![Figure 2.4](image)

**Figure 2.4.** As $x$ approaches $a$, $f(x)$ approaches $L$.

**Example 2.3.** We have $\lim_{x \to 0} 2x \sin x = 0$.

**Solution:** Let $\epsilon$ be any positive real number. Then let $\delta = \epsilon/2$. We know that $|\sin x| \leq 1$ for all $x$. So if $|x - 0| = |x| < \delta = \epsilon/2$, then $|f(x) - 0| = |f(x)| = |2x \sin x| \leq |2x| < 2\delta = \epsilon$, as required.

7.3. One-Sided Limits. There are functions that behave in a certain way up to a point $a$, and then behave very differently after that. We have seen such a function in Example 2.1. The function $g$ of that example satisfied $g(x) = 0$ for negative values of $x$, and $g(x) = 1$ for positive values of $x$. We have seen that $\lim_{x \to 0} g(x)$ does not exist, since no real number $L$ is arbitrarily close to both 0 and 1.

Nevertheless, there are weaker, one-sided notions of limits that are relevant in this example.
Definition 2.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let $a$ be a real number. We say that the left-hand limit of $f$ in $a$ is the real number $L$ if the values of $f(x)$ get arbitrarily close to $L$ and stay arbitrarily close to $L$ when $x$ is suitably close to $a$ and $x < a$.

The fact that $L$ is the left-hand limit of $f$ in $a$ is denoted by

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

For instance, if $g$ is the function defined in Example 2.1, then

$$\lim_{x \to 0^-} g(x) = 0.$$ 

Indeed, if we choose $x$ close to 0 but less than 0, then $g(x) = 0$, so $g(x)$ is arbitrarily close (in fact, equal) to 0.

Definition 2.4. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let $a$ be a real number. We say that the right-hand limit of $f$ in $a$ is the real number $L$ if the values of $f(x)$ get arbitrarily close to $L$ and stay arbitrarily close to $L$ when $x$ is suitably close to $a$ and $x > a$.

The fact that $L$ is the right-hand limit of $f$ in $a$ is denoted by

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

For instance, if $g$ is the function defined in Example 2.1, then

$$\lim_{x \to 0^+} g(x) = 1.$$ 

Indeed, if we choose $x$ close to 0 but more than 0, then $g(x) = 1$, so $g(x)$ is arbitrarily close (in fact, equal) to 1.

At this point, the reader should compare the definitions of limit, left-hand limit, and right-hand limit. The definition of limit (Definition 2.1) imposes the strongest requirements on the values of $f$. Indeed, the values of $f(x)$ have to be close to $L$ when $x$ is close to $a$ and $x < a$ and also when $x$ is close to $a$ and $x > a$. The definitions of the left-hand and right-hand limits impose weaker requirements in that each definition only requires that $f(x)$ be close to $L$ when $x$ is on a given side of $a$ and close to $a$.

It then follows—and the reader should spend a minute verifying it—that if $\lim_{x \to a} f(x) = L$, then $\lim_{x \to a^-} f(x) = L$ and $\lim_{x \to a^+} f(x) = L$.

Conversely, if both the left-hand limit and the right-hand limit of $f$ in $a$ is equal to $L$, then the limit of $f$ in $a$ exists and is equal to $L$.

At this point, the reader should check his or her understanding of the material by considering the function

$$h(x) = \frac{x}{|x|}$$
as \( x \) approaches 0 and deciding if the limits \( \lim_{x \to 0} h(x) \), \( \lim_{x \to 0^-} h(x) \), and \( \lim_{x \to 0^+} h(x) \), exist. It may help to consult Figure 2.5.

![Figure 2.5. Graph of \( h(x) = x/|x| \).](image)

**7.4. Infinite Limits.** In our definitions of limits in this section, the limit \( L \) was always a real number. In this section, we extend those definitions to the cases of *infinite* limits. If \( L = \infty \), then the values of \( f \) have to get arbitrarily close to \( \infty \); that is, they have to get as large as we want. This is the content of the following definition.

**Definition 2.5.** Let \( f : \mathbb{R} \to \mathbb{R} \) be a function. We say that the limit of \( f \) in \( a \) is \( \infty \) if we can get \( f(x) \) arbitrarily large and keep it arbitrarily large if we choose \( x \) suitably close to \( a \) without being equal to \( a \).

Similarly, if \( g : \mathbb{R} \to \mathbb{R} \) is a function, we say that the limit of \( g \) in \( a \) is \( -\infty \) if we can make \( g(x) \) a negative number with an arbitrarily large absolute value and keep \( g(x) \) that way if we choose \( x \) suitably close to \( a \) without being equal to \( a \).

The fact that the limit of \( f \) in \( a \) is \( \infty \) is denoted by

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

**Example 2.4.** Let \( f(x) = 1/x^2 \). Then \( \lim_{x \to 0} f(x) = \infty \).

**Solution:** If we want \( f(x) \) to be larger than an arbitrary positive real number \( N \), all we need to do is to choose \( x \) from the interval \((-\sqrt{1/N}, \sqrt{1/N})\). Then \( x^2 < 1/N \) will hold, implying that \( f(x) = 1/x^2 > N \).

Similarly, if \( g(x) = -1/x^4 \), then \( \lim_{x \to 0} g(x) = -\infty \). Note that if the limit of a function at a given point \( a \) is \( \infty \) or \( -\infty \), then, as \( x \) approaches \( a \), the graph of the function will approach a vertical line intersecting the horizontal axis at \( x = a \). This phenomenon is referred to by saying that \( f \) has a *vertical asymptote* at \( a \).
7.4.1. The Precise Definition of Infinite Limits. The formal definition of infinite limits is similar to that of finite limits. The difference lies in the fact that it is not the same to be close to $\infty$ or to be close to a real number.

**Definition 2.6.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that the limit of $f$ in $a$ is $\infty$ if, for all positive real numbers $N$, there exists $\epsilon > 0$ such that if $|x - a| < \epsilon$, then $f(x) > N$.

Similarly, let $g : \mathbb{R} \to \mathbb{R}$ be a function. We say that the limit of $g$ in $a$ is $-\infty$ if for all negative real numbers $M$, there exists $\epsilon > 0$ such that if $|x - a| < \epsilon$, then $g(x) < M$.

7.4.2. One-Sided Infinite Limits. One-sided infinite limits are defined in an analogous way, as we can see in the following definition.

**Definition 2.7.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let $a$ be a real number. We say that the left-hand limit of $f$ in $a$ is $\infty$ if the values of $f(x)$ get arbitrarily large and stay arbitrarily large when $x$ is suitably close to $a$ and $x < a$.

Similarly, we say that the right-hand limit of $f$ in $a$ is $\infty$ if the values of $f(x)$ get arbitrarily large and stay arbitrarily large when $x$ is suitably close to $a$ and $x > a$.

**Example 2.5.** Let $f(x) = 1/x$. Then $f$ is not defined in 0. Furthermore, $\lim_{x \to 0^-} = -\infty$ and $\lim_{x \to 0^+} = \infty$. As the two one-sided limits are different, $\lim_{x \to 0}$ does not exist.

**Solution:** We can make $f(x) = \frac{1}{x}$ smaller than any given negative number $M$ by choosing $x$ from the interval $(1/M, 0)$. We can make $x$ larger than any positive number $P$ by choosing $x$ from the interval $(0, P)$.

7.5. Exercises.

1. Find $\lim_{x \to -3} x + 7$.
2. Find $\lim_{x \to -13} x^2 - 10x + 7$.
3. Find $\lim_{x \to -3} \frac{x^2 - 4x + 3}{x - 3}$.
4. Find $\lim_{x \to -1} \frac{x^2 + x - 2}{x - 1}$.
5. Does $\lim_{x \to -3} \frac{x^2 - 4x + 7}{x - 3}$ exist?
6. Find $\lim_{x \to 0} \cos x$.
7. Find $\lim_{x \to 0} \frac{x^2}{|x|}$.
8. Find $\lim_{x \to -1} \frac{\sqrt{x + 1}}{x + 1}$.
9. Find $\lim_{x \to -2} \frac{x^3 - 8}{x - 2}$.
(10) Find \( \lim_{x \to -2} \frac{x^3 + 8}{x + 2} \).

(11) Let \( f(x) = \lfloor x \rfloor \) be equal to the largest integer that is at most as large as \( x \). So \( f(3.7) = 3 \). Note that \( f \) is often called the floor function or integer part function. Find the values \( a \) for which \( \lim_{x \to a} f(x) \) exists. If \( a \) is such that \( f \) has no two-sided limit at \( a \), decide if \( f \) has one-sided limits at \( a \).

(12) Let \( g(x) = \lceil x \rceil \) be equal to the smallest integer that is at least as large as \( x \). So \( g(3.7) = 4 \). Note that \( g \) is often called the ceiling function. Find the values \( a \) for which \( \lim_{x \to a} g(x) \) exists. If \( a \) is such that \( g \) has no two-sided limit at \( a \), decide if \( g \) has one-sided limits at \( a \).

(13) Does \( \lim_{x \to \pi/2} \tan x \) exist?

(14) Does \( \lim_{x \to 0^-} \cot x \) exist?

(15) Does \( \lim_{x \to 0^+} \cot x \) exist?

(16) Does \( \lim_{x \to 0} \frac{1}{x} \) exist?

(17) Give an example of a function \( f \) such that \( \lim_{x \to 0^-} f(x) = 0 \), and \( \lim_{x \to 0^+} f(x) = \infty \).

(18) Does \( \lim_{x \to 0} \left( \frac{1}{x^2} + \frac{1}{x} \right) \) exist?

(19) Does \( \lim_{x \to 0} \left( \frac{1}{x^2} + \frac{1}{x} \right) \) exist?

(20) Give an example of a function \( f \) such that \( \lim_{x \to 1^-} f(x) = \infty \), \( \lim_{x \to 1^+} f(x) = -\infty \), and \( f(1) \) is a real number.

8. Limit Laws

8.1. Basic Limit Laws. If \( f \) and \( g \) are two functions and we know the limit of each of them at a given point \( a \), then we can easily compute the limit at \( a \) of their sum, difference, product, constant multiple, and quotient. The rules that provide this limit are given below, and they are very similar to the ways in which the sum, difference, product, constant multiple, and quotient of two functions are defined. Indeed,

(I) \[ \lim_{x \to a} (f + g)(x) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x), \]

(II) \[ \lim_{x \to a} (f - g)(x) = \lim_{x \to a} f(x) - \lim_{x \to a} g(x), \]

(III) \[ \lim_{x \to a} (f \cdot g)(x) = \lim_{x \to a} f(x) \cdot \lim_{x \to a} g(x), \]

(IV) \[ \lim_{x \to a} (c \cdot f)(x) = c \cdot \lim_{x \to a} f(x), \]

where \( c \) is a real number, and
\[ \lim_{x \to a} \left( \frac{f}{g} \right)(x) = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)} \]

if \( \lim_{x \to a} g(x) \neq 0 \).

It is not difficult to believe that these rules are valid. For instance, if \( f(x) \) gets arbitrarily close to \( L \) as \( x \) approaches \( a \) and \( g(x) \) gets arbitrarily close to \( L' \) as \( x \) approaches \( a \), then, as \( x \) approaches \( a \), the value of \( f(x) + g(x) \), that is, the value of \( (f + g)(x) \), will get arbitrarily close to \( L + L' \). This intuitive argument can be made formal using the precise definition of limits.

**Example 2.6.** Let \( f(x) = |x| \) and let \( g(x) = x^2 \). Find the limits of \( f + g \), \( f - g \), \( fg \), \( 3f + 2g \), and \( f/g \) at \( a = 2 \).

**Solution:** Based on the five limit laws given earlier, it makes sense to first compute the limits of \( f \) and \( g \) at 2. The reader is invited to verify that

\[ \lim_{x \to 2} f(x) = \lim_{x \to 2} |x| = \lim_{x \to 2} x = 2, \]

and

\[ \lim_{x \to 2} g(x) = \lim_{x \to 2} x^2 = \lim_{x \to 2} x \cdot \lim_{x \to 2} x = 2 \cdot 2 = 4, \]

where we used the fact that \( g(x) = x^2 = x \cdot x \), so law III can be applied to compute the limit of \( g \) at 2.

Now it is simply a matter of basic algebra to compute the five limits that we have been asked to find. Indeed, applying the five limit laws, we get that

- (I) \( \lim_{x \to 2} (f + g)(x) = \lim_{x \to 2} f(x) + \lim_{x \to 2} g(x) = 2 + 4 = 6 \),
- (II) \( \lim_{x \to 2} (f - g)(x) = \lim_{x \to 2} f(x) - \lim_{x \to 2} g(x) = 2 - 4 = -2 \),
- (III) \( \lim_{x \to 2} (f \cdot g)(x) = \lim_{x \to 2} f(x) \cdot \lim_{x \to 2} g(x) = 2 \cdot 4 = 8 \),
- (IV) \( \lim_{x \to 2} (3f + 2g)(x) = 3 \lim_{x \to 2} f(x) + 2 \lim_{x \to 2} g(x) = 3 \cdot 2 + 2 \cdot 4 = 14 \) (note that here we applied limit law IV to first \( f \), then to \( g \), and then we applied law I to \( 3f \) and \( 2g \)), and
- (V) \( \lim_{x \to 2} \left( \frac{f}{g} \right)(x) = \frac{\lim_{x \to 2} f(x)}{\lim_{x \to 2} g(x)} = \frac{2}{4} = \frac{1}{2}. \)

\[ \square \]

8.2. Frequently Used Special Cases of Limit Laws. A few special cases of limit laws I–V are used so frequently that it is worth mentioning them separately. First, if we repeatedly multiply a function by itself,
we get a power of that function. Applying law III each time, we get that for all positive integers $n$,

$$\lim_{x \to a} (f(x))^n = \left(\lim_{x \to a} f(x)\right)^n.$$  \hfill (2.1)

Note that we have essentially applied this rule in the special case of $n = 2$ when we computed $\lim_{x \to 2} x^2$ in Example 2.6.

The reader is invited to verify that the limits of the constant function $f(x) = c$ and the identity function $f(x) = x$ are given by $\lim_{x \to a} c = c$ for all $a$ and $\lim_{x \to a} x = a$. Formal proofs will be given in the next section.

Applying Equation (2.1) to the identity function $f(x) = x$ yields the equation

$$\lim_{x \to a} x^n = a^n.$$  \hfill (2.2)

It turns out (though it is not obvious) that in Equation (2.1) the exponent $n$ can be replaced by $1/n$; in other words, powers can be replaced by roots, yielding

$$\lim_{x \to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \to a} f(x)}.$$  \hfill (2.3)

(Here $f(x)$ has to be nonnegative if $n$ is even.) So, in particular, if $f(x) = x$, then

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

8.3. Other Useful Facts About Limits. In this section, we discuss a few facts about limits that are often used to compute limits, but are slightly different in nature from the limit laws we discussed so far.

First, let us recall that the definition of $L = \lim_{x \to a} f(x)$ requires that $f(x)$ get arbitrarily close to $L$ if $x$ is sufficiently close to $a$ but not equal to $a$. That is, the value of $f(a)$ does not have to satisfy any requirements. In fact, we can change $f(a)$ to anything we want, and $L = \lim_{x \to a} f(x)$ will not change. What matters is what happens at points other than $a$. Hence, we can conclude that if $f(x) = g(x)$ for all points $x \neq a$, then $\lim_{x \to a} f(x) = \lim_{x \to a} g(x)$ as long as these limits exist. For instance, let $f(x) = (x^2 - 4)/(x - 2)$ for all real numbers $x \neq 2$ and let $f(2) = 2014$. Let $g(x) = x + 2$ for all real numbers. Then $f(x) = g(x)$ unless $x = 2$, and hence $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 4$.

The statement that if $f(x) = g(x)$ for all points $x \neq a$, then $\lim_{x \to a} f(x) = \lim_{x \to a} g(x)$ as long as these limits exist can be significantly strengthened. See Exercise 8.4.1 for a possible direction for that.
Second, Equation (2.2) can be interpreted by saying that the limit of a power function \( f(x) = x^n \) at any point \( a \) is simply the value of \( f(a) \). Now note that polynomials are nothing else but sums of constant multiples of power functions with nonnegative integer exponents. Hence, using limit laws I and IV, we get the following theorem.

**Theorem 2.1.** Let \( p \) be a polynomial function. Then, for any real number \( a \), we have

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

Now recall that a rational function is just the ratio of two polynomials. Hence, using limit law V, we get the following statement from Theorem 2.1.

**Corollary 2.1.** Let \( R(x) \) be a rational function and let \( a \) be a real number such that \( R(a) \) is defined. Then

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

**Proof.** If \( R(x) = p(x)/q(x) \), where \( p \) and \( q \) are polynomials, then by first applying limit law V, and then Theorem 2.1, we get

\[
\lim_{x \to a} R(x) = \lim_{x \to a} \frac{p(x)}{q(x)} = \lim_{x \to a} \frac{p(x)}{q(x)} \lim_{x \to a} \frac{p(x)}{q(x)} = \frac{p(a)}{q(a)} = R(a).
\]

So far all the relationships that we discussed for limits involved equations. We will now discuss two rules that, involve inequalities.

**Theorem 2.2.** Let \( f \) and \( g \) be two functions and assume that, for all real numbers \( x \), the inequality \( f(x) \leq g(x) \) holds. Then

\[
\lim_{x \to a} f(x) \leq \lim_{x \to a} g(x)
\]

for any real number \( a \) as long as both limits exist.

**Proof.** If (2.4) did not hold, then

\[
L_f = \lim_{x \to a} f(x) = D + \lim_{x \to a} g(x) = D + L_g
\]

would hold, for some positive real number \( D \). That would lead to a contradiction, since if \( x \) is so close to \( a \) that \( |f(x) - L_f| < (D/3) \), then, in particular, \( f(x) > L_f - (D/3) \), so

\[
g(x) > L_f - \frac{D}{3} = L_g + \frac{2D}{3}.
\]

This inequality says that no matter how close \( x \) is to \( a \), the distance between \( g(x) \) at \( L_g \) is more than \( 2D/3 \). This contradicts the definition
of \( L_g \), since if \( L_g \) exists, then the values of \( g(x) \) should get arbitrarily close to it, provided that \( x \) is sufficiently close to \( a \).

Note that in Theorem 2.2, the fact that the inequalities are not strict is important. See Exercise 8.4.7 for a relevant question.

**Corollary 2.2 (Squeeze Principle).** If \( f, g, \) and \( h \) are functions such that, for all real numbers \( x \), the inequality \( f(x) \leq g(x) \leq h(x) \) holds and

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

then \( \lim_{x \to a} g(x) \) exists and \( \lim_{x \to a} g(x) = L \).

See Figure 2.6 for an illustration of this important principle.

**Proof.** If \( \lim_{x \to a} g(x) \) exists, then by applying Theorem 2.2 to \( f \) and \( g \), it follows that \( L \leq \lim_{x \to a} g(x) \), and by applying Theorem 2.2 to \( g \) and \( h \), it follows that \( \lim_{x \to a} g(x) \leq L \). So if \( \lim_{x \to a} g(x) \) exists, it is equal to \( L \). In Exercise 8.4.3 you are asked to prove that this limit exists.

The squeeze principle is very useful since it allows us to compute the limits of rather complicated functions as long as we can squeeze them between two functions with identical limits.

![Figure 2.6. Concept of squeeze theorem where \( f(x) \leq g(x) \leq h(x) \).](image)

**Example 2.7.** Let \( g(x) = x \cos(\log x) \). Then \( \lim_{x \to 0} g(x) = 0 \).
Solution: Indeed, let \( f(x) = -x \) and \( h(x) = x \). Then, since \( \cos(\log z) \) is always a real number in the interval \([-1, 1]\), the inequality \( f(x) \leq g(x) \leq h(x) \) holds for all real numbers \( x \). Furthermore, \( \lim_{x \to 0} f(x) = \lim_{x \to 0} h(x) = 0 \), so we can apply Corollary 2.2 to prove our claim.

We could not have used limit law III to compute \( \lim_{x \to 0} g(x) \) since \( \lim_{x \to 0} \cos(\log x) \) does not exist. You are asked to prove this in Exercise 8.4.4.

8.4. Exercises.

1. Find \( \lim_{x \to 2} 3x^2 + 4x + 9 \).
2. Find \( \lim_{x \to 3} \frac{3x^2 + 5x - 2}{x + 1} \).
3. Find \( \lim_{x \to 2} \frac{x^4}{\sqrt{x^2 - 2}} \).
4. Find \( \lim_{x \to 4} \frac{x^2 + 2x + 5}{x^3 + 1} \).
5. Find \( \lim_{x \to 2} \frac{x^2 - 4}{x^3 + 8} \).
6. Let \( f(x) \) and \( g(x) \) be two functions that only differ for a finite number of values of the variable \( x \). Is it true that \( \lim_{x \to a} f(x) = \lim_{x \to a} g(x) \) as long as these limits exist? Why or why not?
7. Find an example of two functions \( f \) and \( g \) such that \( f(x) < g(x) \) for all real numbers \( x \), but there exists a real number \( a \) such that \( \lim_{x \to a} f(x) = \lim_{x \to a} g(x) \).
8. Explain why \( \lim_{x \to a} g(x) \) exists if the conditions of Corollary 2.2 hold.
9. Prove that \( \lim_{x \to 0} \cos(\log x) \) does not exist.
10. Prove that \( \lim_{x \to 0} |x \sin(\sqrt{x})| = 0 \).
11. Compute \( \lim_{x \to 0} x^3 \sin(1/x) \).
12. Compute \( \lim_{x \to 0} \sqrt{x^4 + x^5} \sin(\ln x) \).
13. Compute \( \lim_{x \to 0} \frac{1}{1 + \sqrt{x}} \).
14. Compute \( \lim_{x \to 1} \frac{x^2 + x^5}{x^2 + x^3} \).
15. Compute \( \lim_{x \to 1} \frac{\sqrt{x}}{x^2 + 1} \).
16. Compute \( \lim_{x \to 10} \sqrt[3]{18 - x} \).
17. Let \( a \) be a positive real number. Prove that \( \lim_{x \to a} \frac{\sqrt{x} - \sqrt{a}}{x - a} = \frac{1}{2\sqrt{a}} \).
18. Compute \( \lim_{x \to 3} \frac{x - 3}{x + 1} \).
19. Compute \( \lim_{x \to 4} (x + 4)^{2.5} \). Explain which laws you are using.
20. Compute \( \lim_{x \to 30} (x + 2)^{0.4} \). Explain which laws you are using.
9. Continuous Functions

Intuitively speaking, a function is called continuous at a point \( x = a \) if its graph in a neighborhood of \( x = a \) can be drawn without lifting the pencil from the paper, that is, by a “continuous” line. The formal definition of continuity is as follows.

**Definition 2.8.** A function \( f \) is called continuous at \( a \) if the equality

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

holds.

Note that Definition 2.8 really requires three things. The limit of \( f \) at \( a \) must exist, the function \( f \) must be defined in \( a \) such that \( f(a) \) exists, and the value of \( f(a) \) must agree with the limit of \( f \) at \( a \).

If all these conditions hold, then the behavior of \( f \) at \( a \) is very similar to the behavior of \( f \) around \( a \); in particular, the graph of \( f \) can be drawn without lifting the pencil from the paper. If we had to lift the pencil from the paper, that would mean that some kind of “gap” would exist in the graph of \( f \), so the requirements of Definition 2.8 would not be satisfied.

If a function \( f : \mathbb{R} \to \mathbb{R} \) is continuous at all \( a \in \mathbb{R} \), then it is called continuous. If \( f \) is continuous at each point of the open interval \((c, d)\), then we say that \( f \) is continuous on \((c, d)\). Finally, if you really want a formal definition, the neighborhood of \( a \) is a set \( S \) that contains an open interval \((c, d)\) containing \( a \).

### 9.0.1. The Precise Definition of Continuity.

As the informal definition of continuity is very close to that of limits, it is not surprising that their precise definitions are also similar.

**Definition 2.9.** Let \( f \) be defined in an open interval containing \( a \). We say that \( f \) is continuous in \( a \) if, for all \( \epsilon > 0 \), there exists \( \delta > 0 \) such that if \( |x - a| < \delta \), then \( |f(x) - f(a)| < \epsilon \).

### 9.1. Examples of Continuous Functions.

Let us consider some of the most frequently used continuous functions.

**Example 2.8.** Polynomial functions are continuous.

**Solution:** This is a direct consequence of Theorem 2.1, which we discussed in the last section. Theorem 2.1 stated that the limit of
a polynomial function at \( a \) is equal to the \textit{value} of the polynomial at \( a \), which is precisely what the definition of continuity requires.

There are many classes of functions that are continuous at every point where they are \textit{defined}. If they are not defined somewhere, then, of course, they cannot be continuous there.

\textbf{Example 2.9.} The following are examples of functions that are continuous in every point where they are defined.

(I) \textit{Rational functions}

(II) \textit{Exponential functions}

(III) \textit{Trigonometric functions}

(IV) \textit{Logarithmic functions}

(V) \textit{Inverse trigonometric functions}

The reader is invited to recall the graphs of each of these functions and verify that they consist of continuous lines as long as they are defined.

\textbf{9.2. Functions That Are Not Continuous.} It is time to stop for a moment and think about functions that are not continuous at a given point \( a \). There can be three reasons for this. First, it could be that \( f(a) \) is not defined, for instance, when \( f \) is a rational function whose denominator becomes 0 when \( x = a \). Or it could be that \( g \) is defined at \( a \), but \( \lim_{x \to a} g(x) \) does not exist at \( a \). An example of this is the function defined by \( g(x) = 1 \) if \( x \geq 0 \) and \( g(x) = 0 \) if \( x < 0 \). As we have seen before, the limit of this function does not exist in \( a = 0 \), even if \( g(0) \) is defined. So \( g \) is not continuous at 0. Finally, it could happen that \( h \) is defined in \( a \) and the limit of \( h \) at \( a \) exists, but \( h(a) \) is not equal to this limit. That happens, for example, if \( h(x) = (x + 3)/(x^2 - 9) \) if \( |x| \neq 3 \) and \( h(x) = 1 \) if \( |x| = 3 \). Let \( a = -3 \). Then

\[
h(a) = 1 \neq \lim_{x \to a} h(x) = -\frac{1}{6}.
\]

The interested reader is invited to think about the following example.

\textbf{Excursion 2.1.} \textit{The following function is not continuous anywhere. Let} \( f(x) = 1 \) if \( x \) \textit{is rational} \textit{and let} \( f(x) = 0 \) if \( x \) \textit{is irrational}.

\textbf{9.3. New Continuous Functions from Old.} It follows from the limit laws that several transformations preserve the continuous property of functions.
Theorem 2.3. Let \( f \) and \( g \) be two functions that are continuous at \( a \) and let \( c \) be a real number. Then all of the following are also continuous functions at \( a \):

(I) \( f + g \),
(II) \( f - g \),
(III) \( f \cdot g \),
(IV) \( cf \), and
(V) \( f/g \) as long as \( g(a) \neq 0 \).

Example 2.10. It follows from successive applications of the previous theorem that \( h(x) = e^x \cdot \sin x + 3 \ln x - \sqrt{x} \) is continuous at all positive real numbers \( a \).

The following important theorem also holds, though it is not a direct consequence of our limit laws.

Theorem 2.4. Let \( f \) and \( g \) be two functions such that \( f \) is continuous at \( a \) and \( g \) is continuous at \( f(a) \). Then the composition function \( f \circ g \) is continuous at \( a \).

This theorem is important since it enables us to prove the continuity of functions that would otherwise be cumbersome to handle.

Example 2.11. The function \( h(x) = \sqrt{2 + \sin x} \) is continuous at all real numbers \( a \).

Solution: Let \( f(x) = 2 + \sin x \) and let \( g(x) = \sqrt{x} \). Then \( f \) is continuous everywhere, and \( g \) is continuous at all positive real numbers. As \( f(x) \) is always a positive real number, the statement follows.

9.4. One-Sided Continuity. A function may happen to be continuous in only one direction, either from the “left” or from the “right.” Formally, this means the following.

Definition 2.10. We say that the function \( f \) is left-continuous at \( a \) if \( f(a) = \lim_{x \to a^-} f(x) \). Similarly, we say that \( f \) is right-continuous at \( a \) if \( f(a) = \lim_{x \to a^+} f(x) \).

Example 2.12. Let \( g \) be the function defined by \( g(x) = 1 \) if \( x \geq 0 \) and \( g(x) = 0 \) if \( x < 0 \). Then \( \lim_{x \to 0^-} g(x) = 0 \neq 1 = g(0) \), so \( g \) is not left-continuous at 0. On the other hand, \( \lim_{x \to 0^+} g(x) = 1 = g(0) \), so \( g \) is right-continuous at 0.

The reader is invited to verify that \( f \) is continuous at \( a \) if and only if \( f \) is both left-continuous and right-continuous at \( a \).
We say that a function is continuous on an interval \([a, b]\) if it is continuous at all points of \((a, b)\), left-continuous at \(a\), and right-continuous at \(b\).

**9.5. Intermediate Value Theorem.** Perhaps the most important property of continuous functions is that they do not skip any values between two values that they actually take. For instance, if a tree grows from 3 feet to 6 feet, then there is a time in between when the tree is exactly 4.47 feet tall. The intuitive reason for this is that if there were a value in between that is not taken by the function, then there would be a gap in the graph of the function, contradicting the requirement that the function be continuous. This is the content of the next theorem.

**Theorem 2.5 (Intermediate Value Theorem).** Let \(f\) be a function that is continuous on the interval \([a, b]\). Then, if \(f(a) = y_1\) and \(f(b) = y_2\) and \(y\) is a real number that is between \(y_1\) and \(y_2\), then there exists \(x \in [a, b]\) such that \(f(x) = y\).

In other words, \(f\) takes all values between \(y_1\) and \(y_2\) on the interval \([a, b]\).

**Example 2.13.** There is a real number \(x\) in the interval \([0, 1]\) such that \(x + e^x = 2\).

**Solution:** Let \(f(x) = x + e^x\). Then \(f\) is continuous everywhere, \(f(0) = 1\), and \(f(1) = 1 + e > 3.71\). So, by the intermediate value theorem, we get that \(f\) takes all values between 1 and \(1 + e\) on that interval, including \(y = 2\).

**9.6. Exercises.**

1. Is \(e^{3x+7} \sin x\) continuous everywhere?
2. Is \((x^2 + 1) \ln(x + 1)\) continuous everywhere?
3. Is \(\frac{x^3 + 2x^2 + 3x + 1}{x^2 + 4}\) continuous everywhere?
4. In what point is \(\sqrt{x}\) right-continuous, but not continuous?
5. Where is \(|\ln x|\) continuous?
6. Where is \(\frac{x^2 + 2x + 37}{e^x}\) continuous?
7. Where is \(\tan x\) continuous?
8. Where is \(1/x\) not continuous?
9. Is there a point in which \(x^{-2}\) is left-continuous, but not continuous?
10. Where is \(\frac{1}{1-\ln(1-x)}\) continuous?
11. Where is \(\frac{3x^2}{x^2 + 2}\) continuous?
12. Where is \(\sin(x^2)\) continuous?
(13) Let \( f(x) = x^2 + 3 \) if \( x \neq 2 \). What should \( f(2) \) be if \( f \) is to be a continuous function?

(14) Let \( f(x) = (x^2 - 16)/(x - 4) \) if \( x \neq 4 \). What should \( f(4) \) be if \( f \) is to be a continuous function?

(15) Let \( f(x) = [x] \). Determine the set of points \( a \) for which \( f \) is continuous at \( a \). What can be said about \( f \) at the points where \( f \) is not continuous?

(16) Let \( g(x) = [x] \). Determine the set of points \( a \) for which \( g \) is continuous at \( a \). What can be said about \( g \) at the points where \( g \) is not continuous?

(17) Prove that the equation \( x^5 - x - 1 = 0 \) has a root in the interval \((-1, 2)\).

(18) Prove that the equation \( x^3 - 3x - 1 = 0 \) has at least two roots in the interval \((-1, 2)\).

(19) Prove that the equation \( x^4 + x = \sqrt{2} \) has at least one solution in the interval \((0, 1)\).

(20) Define a function \( f : \mathbb{R} \to \mathbb{R} \) that is not continuous in any point \( a \), and \( f(x) \leq x \) holds for all \( x \in \mathbb{R} \).

10. Limits at Infinity

10.1. Finite Limits at Infinity. In Section 7, we defined what it meant for a function to have a limit \( L \) at a real number \( a \). In this section, we extend that definition and define what it means for a function to have a limit \( L \) at \( \infty \) or at \( -\infty \).

**Definition 2.11.** Let \( f : \mathbb{R} \to \mathbb{R} \) be a function that is defined on some interval \((b, \infty)\). We say that the limit of \( f \) at \( \infty \) is the real number \( L \) if the values of \( f(x) \) get arbitrarily close to \( L \) and stay arbitrarily close to \( L \) when \( x \) is suitably large.

The fact that the limit of \( f \) at \( \infty \) is \( L \) is expressed by the notation

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

This definition follows the idea of the definition of limits at finite points. Indeed, in order for \( \lim_{x \to \infty} f(x) = L \) to hold, we require that the values of \( f(x) \) get arbitrarily close to \( L \) and stay arbitrarily close to \( L \) if \( x \) is large enough. Here “\( x \) is large enough” means that \( x \) is in a suitably selected neighborhood of \( \infty \), in other words, in an open interval \((c, \infty)\). Recall that this is analogous to what we required in the finite case. There we said that \( \lim_{x \to a} f(x) = L \) if \( f(x) \) got arbitrarily close to \( L \) and stayed arbitrarily close to \( L \) once \( x \) was suitably close to \( a \), that is, when \( x \) was in a suitably selected neighborhood of \( a \).
Example 2.14. Let \( f(x) = 1/x \). Then
\[
\lim_{x \to \infty} f(x) = 0.
\]

Solution: If we want the value of \( f(x) \) to be closer than \( \epsilon \) to 0, all we have to do is to select \( x \) such that \( x > 1/\epsilon \) holds. Once \( x \) gets past \( 1/\epsilon \), the values of \( f(x) \) will stay between 0 and \( \epsilon \). □

The definition of limits at \(-\infty\) is what the reader probably expects.

Definition 2.12. Let \( f : \mathbb{R} \to \mathbb{R} \) be a function defined on some interval \((\infty, b)\). We say that the limit of \( f \) at \(-\infty\) is the real number \( L \) if the values of \( f(x) \) get arbitrarily close to \( L \) and stay arbitrarily close to \( L \) when \( x \) is a negative number with a suitably large absolute value.

The fact that the limit of \( f \) at \(-\infty\) is \( L \) is expressed by the notation
\[
\lim_{x \to -\infty} f(x) = L.
\]

Example 2.15. Let \( f(x) = 1/x^2 \). Then
\[
\lim_{x \to -\infty} f(x) = 0.
\]

Solution: If we want to get \( f(x) \) closer than \( \epsilon \) to 0 and keep it there, it suffices to choose \( x \) such that \( x < -1/\sqrt{\epsilon} \). Then \( x^2 > 1/\epsilon \), and hence \( f(x) = 1/x^2 < \epsilon \). □

10.1.1. The Formal Definition of Limits at Infinity. The formal definition of limits at infinity is very similar to that of limits at finite points. The only difference is in the formal description of what it means to be in a neighborhood of infinity versus what it means to be in a neighborhood of a real number.

Definition 2.13. Let \( f : \mathbb{R} \to \mathbb{R} \) be a function defined on some interval \((b, \infty)\). We say that \( \lim_{x \to \infty} f(x) = L \) if, for all positive real numbers \( \epsilon \), there exists a positive real number \( N \) such that if \( x > N \), then \( |f(x) - L| < \epsilon \).

The formal definition of limits at negative infinity is analogous. The only difference is again in the formal description of what it means for \( x \) to be in a neighborhood of \(-\infty\). It means to be in an interval \((-\infty, c)\).

Definition 2.14. Let \( f : \mathbb{R} \to \mathbb{R} \) be a function defined on some interval \((-\infty, b)\). We say that \( \lim_{x \to -\infty} f(x) = L \) if, for all positive real numbers \( \epsilon \), there exists a negative real number \( N \) such that if \( x < N \), then \( |f(x) - L| < \epsilon \).
10. LIMITS AT INFINITY

10.1.2. The Graphical Meaning of a Finite Limit at Infinity. If a function $f$ has limit $L$ at $\infty$ or $-\infty$, then the graph of the function will approach the horizontal line $y = L$ at that infinity. The graph may or may not actually touch that line or even become that line. The line $y = L$ is called a horizontal asymptote of the graph of $y = f(x)$ when $\lim_{x \to \infty} f(x) = L$ or $\lim_{x \to -\infty} f(x) = L$ holds.

10.2. Infinite Limits at Infinity. It can happen that the limit of a function at $\infty$ is not a real number but rather $\infty$ or $-\infty$.

**Definition 2.15.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined on some interval $(b, \infty)$. We say that the limit of $f$ at $\infty$ is $\infty$, denoted by $\lim_{x \to \infty} f(x) = \infty$, if $f(x)$ gets arbitrarily large and stays arbitrarily large if $x$ gets sufficiently large.

**Example 2.16.** Let $f(x) = e^x$. Then $\lim_{x \to \infty} f(x) = \infty$.

**Solution:** In order to get $f(x)$ to be larger than some given positive real number $M$, it suffices to choose $x > \ln M$.

The following notation is defined in an analogous way:

- (I) $\lim_{x \to \infty} f(x) = -\infty$.
- (II) $\lim_{x \to -\infty} g(x) = \infty$.
- (III) $\lim_{x \to -\infty} h(x) = -\infty$.

Each of these definitions refers to a fact that the values of a function get arbitrarily far away from 0 and stay arbitrarily far away from 0 (in the appropriate direction) if $x$ gets sufficiently far away from 0 (in the appropriate direction). The reader should test his or her understanding of these concepts by verifying that $\lim_{x \to -\infty} 1 - x = -\infty$, while $\lim_{x \to -\infty} x^2 = \infty$, and $\lim_{x \to -\infty} x^3 = -\infty$.

10.2.1. The Formal Definition of Infinite Limits at Infinity. By now, the formal definition of infinite limits at infinity probably does not come as a surprise. We are providing a formal definition for one of the four possible scenarios that can occur due to changes in sign. The other three cases are analogous.

**Definition 2.16.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined on some interval $(b, \infty)$. We say that $\lim_{x \to \infty} f(x) = \infty$ if, for all positive real numbers $M$, there exists a positive real number $N$ such that if $x > N$, then $f(x) > M$. 

10.3. Computing Limits at Infinity. The limit laws that we learned for limits at finite points stay true for limits at infinity as well, provided, of course, that they make sense. Here are a few examples.

Example 2.17. We have
\[ \lim_{x \to \infty} \frac{x + 3}{x - 4} = 1. \]

It would be wrong to argue as follows: “The numerator is the function \( f(x) = x + 3 \), and the denominator is the function \( g(x) = x - 4 \). At \( \infty \), they both have limit \( \infty \), so, by the limit law for quotients, the limit of their quotient is 1.”

The problem with this argument is that \( \infty \) is not a number. So \( \infty / \infty \) is not defined. It is possible for \( f \) and \( g \) both to have limit \( \infty \) at \( \infty \), and for \( f/g \) to have limits \( c \) at \( \infty \), for any given real number \( c \). Indeed, let \( f(x) = cx \) and let \( g(x) = x \).

Instead, we can solve Example 2.17 as follows.

Solution:
\[
\lim_{x \to \infty} \frac{x + 3}{x - 4} = \lim_{x \to \infty} \frac{(x - 4) + 7}{x - 4} \\
= \lim_{x \to \infty} \left( 1 + \frac{7}{x - 4} \right) \\
= 1 + \lim_{x \to \infty} \frac{7}{x - 4} \\
= 1 + 0 \\
= 1.
\]

We would like to point out other pitfalls when dealing with the application of limit laws and infinite limits. The following expressions are not defined:

(I) \( \infty + (-\infty) \)

(II) \( \infty \cdot 0 \) and \( -\infty \cdot 0 \)

(III) \( 1^\infty \) and \( 1^{-\infty} \)

The following theorem is very useful when dealing with limits at \( \infty \).

Theorem 2.6. Let \( r \) be a positive rational number. Then
\[ \lim_{x \to \infty} \frac{1}{x^r} = 0. \]

If \( r \) is an integer, then this statement follows from the fact that \( \lim_{x \to \infty} 1/x = 0 \) by applying limit law III (for products) \( r \) times. If \( r = p/q \), where \( p \) and \( q \) are positive integers, then we can first prove the theorem for \( x^p \), and then, using the root law, for \( x^{p/q} = \sqrt[q]{x^p} \).
Many limits can be computed with the help of this theorem.

**Example 2.18.** We have

\[
\lim_{x \to \infty} \frac{x^2 + 3x + 1}{x^3} = 0.
\]

**Solution:** We have

\[
\frac{x^2 + 3x + 1}{x^3} = \frac{x^2}{x^3} + \frac{3x}{x^3} + \frac{1}{x^3},
\]

and each of the three summands has limit 0 at \(\infty\) by the preceding theorem. Hence, by the limit law for sums, so does their sum. \(\Box\)

Note that the limit would not change if we changed the denominator from \(x^3\) to \(x^3 + 3x^2 + 4x + 5\). This would have decreased the value of our function, but would have still kept it positive. Hence, by the squeeze principle, we can then conclude that

\[
\lim_{x \to \infty} \frac{x^2 + 3x + 1}{x^3 + 3x^2 + 4x + 5} = 0.
\]

### 10.4. Exercises.

1. Find \(\lim_{x \to \infty} \frac{x + 1}{x^2 + 4}\).
2. Find \(\lim_{x \to \infty} \frac{3x^2 + 4x + 1}{x^2 + 5}\).
3. Find \(\lim_{x \to \infty} \frac{x^3 + 2x}{x^3 + 4x + 6}\).
4. Find \(\lim_{x \to \infty} \frac{3x^2 + 4x + 1}{x^2 - 4}\).
5. Compute \(\lim_{x \to \infty} \frac{x^5 - 3x}{x^3 - 14}\).
6. Let \(R(x) = p(x)/q(x)\) be a rational function. Explain how \(\lim_{x \to \infty} R(x)\) depends on \(p(x)\) and \(q(x)\).
7. Compute \(\lim_{x \to \infty} \frac{x + 2}{x - 3} + \frac{2x^2 - 3x + 1}{x^2 - 2x + 1}\).
8. Compute \(\lim_{x \to \infty} \frac{\sin(x)}{x}\).
9. Does \(\lim_{x \to \infty} \frac{x}{\sin(x)}\) exist?
10. Compute \(\lim_{x \to \infty} \frac{\sin(x) + \cos^2(x)}{x^3 + 1}\).
11. Compute \(\lim_{x \to \infty} \frac{\sqrt{x^2 + 1}}{x}\).
12. Compute \(\lim_{x \to \infty} \frac{\sqrt{x^2 + 1}}{\sqrt{x^2 - 1}}\).
13. Compute \(\lim_{x \to \infty} \frac{\sqrt{x^2 + 1}}{2x + 11}\).
14. Compute \(\lim_{x \to \infty} \frac{100x + 4}{x^2 - 4x + 5}\).
15. Is there a real number \(L\) such that \(L = \lim_{x \to \infty} \frac{x^3 + 1}{1000x^2 + 9x + 35}\) holds?
(16) Compute \( \lim_{x \to \infty} x^{-0.1} + x^{-0.9} \).

(17) Compute \( \lim_{x \to \infty} \frac{x^2 + x^3 + x^4}{x+1} \).

(18) Compute \( \lim_{x \to \infty} \frac{-x + \sin x}{x} \).

(19) Does \( \lim_{x \to \infty} x \sin x \) exist?

(20) Give an example of a function \( f \) such that \( \lim_{x \to \infty} f(x) = 0 \),
but \( \lim_{x \to \infty} (1/f(x)) \) does not exist.

11. Derivatives

11.1. Tangent Lines. Let us consider a function, such as \( f(x) = x^2 \), and
its graph. Let us choose a point on the graph, say the point \( P = (3, 9) \). Now let us look for the slope of the tangent line to the graph at that point.

That is, consider a sequence of points \( P_1, P_2, \ldots \) that are all on the
graph of \( f \) and are closer and closer to \( P \). For each of these points,
draw the line \( P_i P \). The slope of these lines will approach a certain
slope, and so the lines \( P_i P \) will approach a certain line. That line is
called the tangent line of \( f \) at \( P \). See Figure 2.7 for an illustration.

**Definition 2.17.** Let \( f \) be a function and let \( P = (a, f(a)) \) be a
point on the graph of \( f \). Then the tangent line to \( f \) at \( P \) is the line
that contains \( P \) and has slope

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

provided that this limit exists.

Figure 2.7. Notice that as \( x \) approaches \( a \) the secant
line approaches the tangent line.
The interactive website http://www.math.ufl.edu/~mathguy/ufcalc
book/derivative_def.html provides further examples of this
phenomenon.

Note that in the preceding definition, \((f(x) - f(a))/(x-a)\) is simply
the slope of the line connecting the points \(P\) and \((x, f(x))\).

**Example 2.19.** In our running example, that is, when \(f(x) = x^2\)
and \(P = (3, 9)\), the tangent line is the line that goes through \(P\) and has
slope
\[
\lim_{x \to 3} \frac{f(x) - f(3)}{x - 3} = \lim_{x \to 3} \frac{x^2 - 9}{x - 3} = \lim_{x \to 3} (x + 3) = 6.
\]

### 11.2. Velocities.

Recall that in Section 6, we mentioned that the average velocity of a moving object, such as a car, can be computed by
the rule \(v = s/t\). That is, the average velocity is equal to the distance
covered divided by the time needed to cover that distance. However,
what can be said about the instantaneous velocity, that is, the velocity
in a given moment?

We could not answer that question in Section 6 since we did not
have the tools to handle the fact that when only a given moment is
considered, both the numerator and the denominator of the formula
\(v = s/t\) are 0. Now that we have learned about limits, we can overcome
that difficulty as follows.

**Definition 2.18.** Let \(f(t)\) be a function such that \(f(t)\) is the dis-
tance covered by a moving object in \(t\) units of time. Then the instant-
aneous velocity of the object \(a\) units of time after it starts moving is

\[
v(t) = \lim_{t \to a} \frac{f(t) - f(a)}{t - a},
\]

provided that this limit exists.

**Example 2.20.** A car starts out by accelerating for 10 seconds so
that the distance covered in the first \(t\) seconds is obtained (in meters)
by the function \(f(t) = \frac{1}{2}t^2\) if \(t \leq 10\). What is the instantaneous velocity
of the car after 4 seconds?

**Solution:** By the definition of instantaneous velocity, we must compute

\[
v(4) = \lim_{t \to 4} \frac{f(t) - f(4)}{t - 4} = \lim_{t \to 4} \frac{t^2 - 16}{2(t - 4)} = \lim_{t \to 4} \frac{t + 4}{2} = 4.
\]

So, at the end of the fourth second (exactly 4 seconds after starting
out), the car will move at a rate of 4 meters per second. \(\square\)
11.3. The Derivative of a Function. The fact that the last two concepts, the tangent line and the instantaneous velocity, led to very similar definitions suggests that there is a very general principle at work and we have seen two special cases of that principle.

This is indeed the case.

Definition 2.19. Let $f$ be a function. The derivative of $f$ at $a$ is the limit

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

if this limit exists and is finite.

So, in particular, $f'(a)$ is the slope of the tangent line of $f$ at $a$ (unless that tangent line is vertical). Furthermore, the instantaneous velocity at time $a$ is the derivative of the distance covered (as a function of the time $t$ needed to cover that distance) at $t = a$.

In other words, the derivative is a common generalization of the concepts of tangent line and instantaneous velocity.

11.4. Exercises.

1. Find the slope of the tangent line to the curve $f(x) = 3x^2 - 7$ at the point $(2,5)$.
2. Find the slope of the tangent line to the curve $f(x) = x^3$ at $x = 0$.
3. Find the slope of the tangent line to the curve $f(x) = x(1-x)$ at $x = 1/2$.
4. Find the slope of the tangent line to the curve $f(x) = x^2$ at three different points. Do you see a pattern?
5. Find the slope of the tangent line to the curve $f(x) = x^2 + x$ at three different points. Do you see a pattern?
6. Show an example of a curve that does not have a tangent line at some point $a$ because the limit defined in (2.5) does not exist or is infinite.
7. The distance covered by a car in a certain time period is described by the function

$$f(t) = tm + \frac{t^2(b-m)}{2},$$

where $b$ and $m$ are positive constants. Let us assume that $t \in [0,1]$. Find the instantaneous velocity of the car at a given moment $t = a$.
8. A ball is rolling down a hill. The distance it covers in time $t$ is given by the function $s(t) = 3t + 0.5t^2$, where $t \in [0,5]$
and time is measured in seconds. What is the instantaneous velocity of the ball at the moment of time $t = 3$?

(9) At the beginning of a daily training session, the distance covered by a runner is described by the function $s(t) = 0.5t^2$, where $t \in [0, 5]$, time is measured in seconds, and distance is measured in meters. At what moment will the runner have an instantaneous velocity of 6 m/s?

(10) A car is moving at a speed of 20 meters per second when its driver applies the brakes and the car starts slowing down. The car stops 10 seconds later. The distance covered by the car in $t$ seconds, starting at the moment when the driver steps on the brakes, is given by the function $f(t) = 20t - t^2$ for $t \in [0, 10]$. What is the velocity of the car $t$ seconds after the brakes are applied?

(11) Prove that, for any constant $c$, the derivative of the function $f(x) = c$ is 0 at any point $a$.

(12) Find the derivative of the function $f(x) = x + 5$ at $a = 7$.

(13) Find the derivative of the function $f(x) = 3x + 2$ at $a = 4$. What happens if we change the value of $a$?

(14) Find the derivative of the function $f(x) = 3x - 11$ at $a = 4$. Compare your result with the result of the previous exercise.

(15) Find the derivative of the function $f(x) = 2x^2$ at $a = 2$.

(16) Find the derivative of the functions $g(x) = 2x^2 + 1$ at $a = 2$ and $h(x) = 2(x - 1)^2$ at $a = 3$.

(17) Find the derivative of the function $f(x) = x^3$ at $x = 1$.

(18) Find the derivative of the function $f(x) = \sqrt{x}$ at $x = 4$.

(19) Let

$$g(x) = \begin{cases} 
2x & \text{if } 0 \leq x, \\
x & \text{if } x < 0.
\end{cases}$$

Does $f'(0)$ exist?

(20) Let $f$ be defined as in the previous exercise. Does $f'(a)$ exist if $a \neq 0$?

12. The Derivative as a Function

12.1. Rates of Change. In the last section, we saw that the derivative of a function at a given point was a common generalization of the concepts of tangent lines and instantaneous velocities. We will now further elaborate on that, in order to understand how far-reaching the concept of derivatives is.

If $f$ is a function and $f(x) = y$, then the quantity denoted by $y$ depends on the quantity denoted by $x$. This is sometimes expressed
by saying that \( x \) is the independent variable and \( y \) is the dependent variable. If \( x \) changes, then the change in \( y \) can be described in terms of the change in \( x \).

In particular, if \( x \) changes from \( x_1 \) to \( x_2 \), then \( y = f(x) \) changes from \( y_1 = f(x_1) \) to \( y_2 = f(x_2) \). The average rate of change for the interval \((x_1, x_2)\) is then the ratio

\[
\frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x},
\]

where \( \Delta x \) is the change (or increment) of \( x \). We have to use the word “average” since we only have information about the values of \( y \) at the endpoints of the interval \((x_1, x_2)\); we do not know how \( f(x) = y \) behaves in the rest of the interval. If we want more precise information, such as the instantaneous rate of change of \( f(x) = y \) at a given point, then we have to use the notion of limits again, just as we have done twice in the last section. That is, at a given point \( x = a \), we define the instantaneous rate of change of \( f(x) = y \) as

\[
\lim_{x_2 \to a} \frac{f(x_2) - f(a)}{x_2 - a} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.
\]

**12.2. The Derivative of the Function \( f \).** Recall that, at a given point \( a \), the derivative of the function \( f \) is defined as the limit

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

Note that this definition associates the real number \( f'(a) \) to the real number \( a \). That is, \( f' : \mathbb{R} \to \mathbb{R} \) is a function. The function \( f' \) is called the derivative of \( f \). The operation that takes \( f \) into \( f' \) is called differentiation. This explains the following definition.

**Definition 2.20.** A function \( f \) is called differentiable at \( a \) if \( f'(a) \) exists.

We say that \( f \) is differentiable on the interval \((a, b)\) if \( f \) is differentiable at \( d \) for all \( d \in (a, b) \).

**Example 2.21.** The function \( f(x) = x^3 \) is differentiable in every real number \( a \), and \( f'(a) = 3a^2 \).
Solution: We have
\[
\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{x^3 - a^3}{x - a} = \lim_{x \to a} \frac{(x - a)(x^2 + xa + a^2)}{x - a} = \lim_{x \to a} (x^2 + xa + a^2) = 3a^2.
\]

The functions we have considered so far had only one independent variable, usually the variable \( x \). The dependent variable was usually denoted by \( y \), so \( y = f(x) \) held. So it was always clear that the derivative was taken with respect to \( x \). However, there are circumstances when this is not so clear, usually when \( f \) depends on more than one variable. Therefore, there are additional ways to denote the function \( f' \) such as
\[
\begin{align*}
&\frac{dy}{dx}, \\
&\frac{df}{dx}, \\
&\frac{dx}{dx} f(x), \\
&D_x f(x), \\
&D f(x).
\end{align*}
\]

12.3. Differentiability Versus Continuity. The definitions of differentiability and continuity are similar. Which one imposes stronger requirements on a function at a given point? The following theorem shows that differentiability is the stronger requirement.

**Theorem 2.7.** If \( f \) is differentiable at \( a \), then \( f \) is continuous at \( a \).

**Proof.** If \( f \) is differentiable at \( a \), then

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

in particular, the limit shown on the right-hand side exists. Multiplying both sides by the function \( g(x) = x - a \), we get

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

Now, taking limits at \( a \) on both sides, we obtain

\[
(2.6) \quad f'(a) \cdot \lim_{x \to a} (x - a) = \lim_{x \to a} (f(x) - f(a)),
\]
since we can apply the limit law for products on the right-hand side to get that
\[
\lim_{x \to a} (x - a) \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \left( (x - a) \cdot \frac{f(x) - f(a)}{x - a} \right)
= \lim_{x \to a} (f(x) - f(a)) .
\]
Finally, note that the left-hand side of (2.6) is equal to 0 since \( f'(a) (x - a) \) is a polynomial that takes value 0 when \( x = a \). Hence, the right-hand side of (2.6) is equal to 0 as well, that is,
\[
0 = \lim_{x \to a} (f(x) - f(a)) = (\lim_{x \to a} f(x)) - f(a) .
\]
Adding \( f(a) \) to both the far left and far right sides, we get that
\[
f(a) = \lim_{x \to a} f(x) ,
\]
which means that \( f \) is continuous at \( a \). \( \square \)

The converse of Theorem 2.7 is not true. Indeed, the function \( f(x) = |x| \) is continuous at \( a = 0 \), but it is not differentiable. The reader is invited to prove this by showing that
\[
\lim_{x \to 0^-} \frac{|x| - 0}{x - 0} = \lim_{x \to 0^+} \frac{|x|}{x} \neq \lim_{x \to 0^+} \frac{|x|}{x} = \lim_{x \to 0^-} \frac{|x| - 0}{x - 0} ,
\]
and hence
\[
f'(0) = \lim_{x \to 0} \frac{|x| - 0}{x - 0}
\]
does not exist.

In general, there are several reasons a continuous function may fail to be differentiable at a given point. It could be that the graph of the function has a “corner,” like that of \( |x| \) at 0, and hence the slope of the tangent line cannot be defined because the left-hand limit and the right-hand limit of the lines approaching the purported tangent line are not equal. Or, it could be that the function has a vertical tangent line at the given point. See Exercise 12.5.6 for an example of this.

12.4. Higher-Order Derivatives. In upcoming chapters, it will often be useful to consider not only the derivative of a function but also the derivative of the derivative and even the derivative of the derivative of the derivative. These functions appear so often that they have their own names.

If \( f \) is a differentiable function on an interval \( (a, b) \) and its derivative \( f' \) is also differentiable on \( (a, b) \), then the derivative of \( f' \) is called the second derivative of \( f \) and is denoted by \( f'' \). Similarly, if \( f'' \) is
differentiable on \((a, b)\), then its derivative is called the \textit{third derivative}\ of \(f\) and is denoted by \(f'''\). Higher-order derivatives are defined in an analogous way, but denoted slightly differently. For instance, the seventh derivative of \(f\) is denoted by \(f^{(7)}\), and, in general, the \(n\)th derivative is denoted by \(f^{(n)}\).

\textbf{Example 2.22.} We have seen in Example 2.21 that if \(f(x) = x^3\), then \(f'(x) = 3x^2\). Therefore,

\[
f''(a) = \lim_{x \to a} \frac{f'(x) - f'(a)}{x - a}
= \lim_{x \to a} \frac{3(x^2 - a^2)}{x - a}
= \lim_{x \to a} 3(x + a)
= 6a.
\]

So \(f''(x) = 6x\).

In Exercise 12.5.2, you are asked to prove that \(f'''(x) = 6\) for all \(x\), and in Exercise 12.5.3, you are asked to compute higher-order derivatives of \(f\).

\textbf{12.5. Exercises.}

1. Let \(f(x) = cx + d\), where \(c\) and \(d\) are fixed real numbers. Compute \(f'(x)\), and \(f''(x)\).
2. Let \(f(x) = px^2 + qx + r\), where \(p\), \(q\), and \(r\) are fixed real numbers. Compute \(f'(x)\), \(f''(x)\), and \(f'''(x)\).
3. Let \(f(x) = \sqrt{x}\). Compute \(f'(a)\) at some point \(a > 0\).
4. Compute \(f'(a)\) if \(f(x) = \sqrt{4x + 1}\).
5. Let \(f(x) = x^3\). Prove that \(f'''(x) = 6\) for all real numbers \(x\).
6. Let \(f(x) = x^3\). Compute \(f^{(4)}(x)\). What can be said about higher-order derivatives of \(f\)?
7. Let \(f(x) = x^4\). Compute \(f'(a)\) at some point \(a\).
8. Let \(f(x) = \frac{1}{x}\). Compute \(f'(a)\) at some point \(a \neq 0\).
9. Let \(f(x) = \frac{1}{x^2}\). Compute \(f'(a)\) at some point \(a \neq 0\).
10. Let \(f\) and \(g\) be two functions such that \(f(x) - g(x) = c\) for all \(x\), where \(c\) is a constant. Is it true that, at every point \(a\) where \(f'(a)\) exists, \(g'(a)\) also exists, and \(f'(a) = g'(a)\) holds?
11. Let \(f\) be defined on the interval \([0, 2]\) by \(f(x) = \sqrt{1 - x^2}\) if \(0 \leq x \leq 1\), and \(f(x) = -\sqrt{1 - (x - 2)^2}\) if \(1 < x \leq 2\). So the graph of \(f(x)\) is the union of two quarters of a unit circle. Prove that, at \(x = 1\), the graph of \(f\) has a vertical tangent.
line, that is,
\[
\lim_{x \to 1} \left| \frac{f(x) - f(1)}{x - 1} \right| = \infty.
\]

(12) Find an example of a function \( f \) and a real number \( a \) such that \( f'(a) \) exists, but \( f''(a) \) does not exist.

(13) Find an example of a function \( f \) and a point \( a \) such that \( f'(a) \) and \( f''(a) \) exist, but \( f'''(a) \) does not exist.

In the remaining exercises of this section, decide whether the derivative of the given function in the given point exists or not.

(14) \( f(x) = |x| x \) at \( a = 0 \).

(15) \( f(x) = |x|^2 \) at \( a = 0 \).

(16) \( f(x) = |x| \) at \( a = 1 \).

(17) \( f(x) = \lceil x \rceil \) at \( a = 2 \).

(18) \( f(x) = \ln x \) at \( a = 0 \).

(19) \( f(x) = \cot x \) at \( a = 0 \).

(20) \( f(x) = 2x^2 + x^3 \) at any real number \( a \).
13. Derivatives of Polynomial and Exponential Functions

13.1. Polynomials. Let us recall that polynomials are sums of power functions with nonnegative integer exponents, such as the function \( f(x) = 3x^2 + 4x + 6 \). In this section, we will deduce general rules for the derivatives of polynomial functions. We start by their “building blocks,” power functions. The simplest of these is the class of constant functions.

**Theorem 3.1.** Let \( c \) be a real number and let \( f(x) = c \) for all \( x \). Then \( f'(a) = 0 \) for all real numbers \( a \).

Before we prove the theorem, we point out that, intuitively, it makes perfect sense. The derivative of a function \( f \) describes the rate of change of \( f \), but if \( f \) is a constant function, then \( f \) never changes (it has zero change).

**Proof of Theorem 3.1.** We have

\[
 f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \\
 = \lim_{x \to a} \frac{c - c}{x - a} \\
 = 0.
\]

Note that \( \lim_{x \to a} (c - c)/(x - a) = 0 \) since \( (c - c)/(x - a) = 0 \) for all values \( x \neq a \). \( \square \)

We now turn our attention to a more general class of power functions, those of the form \( f(x) = x^n \), where \( n \) is a positive integer. Let us recall the algebraic identity

\[
x^n - a^n = (x - a) \cdot (x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}).\]

**Theorem 3.2.** Let \( n \) be a positive integer and let \( f(x) = x^n \). Then

\[
f'(a) = a^{n-1}.
\]
PROOF. We have
\[ f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \]
\[ = \lim_{x \to a} \frac{x^n - a^n}{x - a} \]
\[ = \lim_{x \to a} \left( (x - a) \cdot (x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}) \right) \]
\[ = \lim_{x \to a} \left( x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1} \right) \]
\[ = na^{n-1}. \]

Note that this agrees with our result from the last section that showed that if \( f(x) = x^3 \), then \( f'(x) = 3x^2 \).

It turns out that Theorem 3.2 holds even if \( n \) is not a positive integer. That is, for all real numbers \( \alpha \), if \( f(x) = x^\alpha \), then \( f'(x) = \alpha x^{\alpha - 1} \). We will see a formal proof of this fact later. In the exercises, you are asked to prove two special cases of this general result.

13.1.1. Three Simple Rules. Derivatives are limits of certain functions, so it is not surprising that some of the laws governing their computation are very similar to limit laws. That is, if we know the derivative of \( f \) and \( g \), then we can easily compute the derivative of \( f + g \), \( f - g \), and \( cf \), where \( c \) is a given real number. The rules are as follows.

**Theorem 3.3.** Let \( f \) and \( g \) be two functions that are differentiable at \( a \). Then \( f + g \) is differentiable at \( a \), and
\[ (f + g)'(a) = f'(a) + g'(a). \]

**Proof.** We have
\[ (f + g)'(a) = \lim_{x \to a} \frac{(f + g)(x) - (f + g)(a)}{x - a} \]
\[ = \lim_{x \to a} \left( \frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a} \right) \]
\[ = f'(a) + g'(a). \]

The other two rules and their proofs are so similar that they are left as exercises.

**Theorem 3.4.** Let \( f \) and \( g \) be two functions that are differentiable at \( a \). Then \( f - g \) is differentiable at \( a \), and
\[ (f - g)'(a) = f'(a) - g'(a). \]
Theorem 3.5. Let \( f \) be a function that is differentiable at \( a \) and let \( c \) be a real number. Then \( cf \) is differentiable at \( a \) and
\[
(cf)'(a) = cf'(a).
\]

It is very important to point out that the other limit laws do not carry over to derivatives in the same fashion. That is, in general, \((fg)' \neq f'g'\), and \((f/g)' \neq f'/g'\). We will learn some more complicated rules to compute the derivatives of \( fg \) and \( f/g \) in the next section.

Theorems 3.3 to 3.5 enable us to compute the derivative of any polynomial function.

Example 3.1. Let \( p(x) = 3x^3 + 5x^2 - 6x + 8 \). Find \( p'(x) \).

Solution: Note that \( p(x) \) is just a sum (and difference) of constant multiples of power functions. The derivatives of power functions are computed in Theorem 3.2. Then we can apply Theorems 3.3 to 3.5 to get
\[
p'(x) = (3x^3)' + (5x^2)' - (6x)' + (8)' \\
= 3(x^3)' + 5(x^2)' - 6(x)' + (8)' \\
= 9x^2 + 10x - 6.
\]

\[\square\]

13.2. Exponential Functions. Let us now compute the derivative of the exponential function \( f(x) = b^x \), where \( b \) is some positive constant. By the definition of derivatives, we get
\[
f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \\
= \lim_{x \to a} \frac{b^x - b^a}{x - a} \\
= \lim_{z \to 0} \frac{b^{a+z} - b^a}{z} \\
= b^a \lim_{z \to 0} \frac{b^z - 1}{z} \\
= b^a f'(0).
\]

Several comments are in order. First, note the substitution \( z = x - a \) in the third line. Second, note that \( b^a \) is a constant that does not depend on \( z \); hence, the limit law for constant multiples was used in the fourth line. Third, in the special case when \( a = 0 \), the definition of the derivative yields \( f'(0) = \lim_{z \to 0} (b^z - 1)/z \). We used this fact in the last line.
In other words,

\[(3.1) \quad f'(x) = f'(0)b^x = f'(0)f(x).\]

That is, the derivative of the function \(f\) is a constant multiple of \(f\). The constant in question is \(f'(0)\), that is, \(\lim_{z \to 0} (b^z - 1)/z\). Numerical experimentation suggests that the larger \(b\) is, the larger this limit is. Graphical experimentation suggests this as well. Indeed, \(f'(0)\) is the slope of the tangent line to the curve of \(f(x) = b^x\) at the point \(x = 0\), and plotting \(f\) for various values of \(b\) suggests that the larger \(b\) is, the larger this slope is.

In particular, it can be proved that there exists a real number \(e\), close to 2.71, such that

\[\lim_{z \to 0} e^z - 1 = 1.\]

This real number \(e\) is the basis of the natural logarithm that we denote by \(\ln\).

The reader may wish to consult the interactive website http://www.math.ufl.edu/~mathguy/ufcalcbook/exponent.html for further illustrations.

**Definition 3.1.** Let \(e\) be the real number such that

\[\lim_{z \to 0} e^z - 1 = 1.\]

So, in the special case of \(b = e\), Equation (3.1) takes the form

\[(e^x)' = e^x,\]

since

\[f'(0) = \lim_{z \to 0} e^z - 1 = 1.\]

That is, the derivative of \(f(x) = e^x\) is \(f(x) = e^x\) itself. In Section 16, we will see what that implies for the derivatives of exponential functions with bases different from \(e\).

**13.3. Exercises.**

1. Let \(f(x) = x^3 + 2x^2 + 3x + 4\). Compute \(f'(x)\) and \(f''(x)\).
2. Let \(f(x) = x^4 - 3x + 9\). Compute \(f'(x)\) and \(f''(x)\).
3. Let \(f(x) = x^8 - 2x^4 + 1\). Compute \(f'(x)\), \(f''(x)\), and \(f'''(x)\).
4. Let \(f(x) = -x + x^3\). Compute all derivatives (first, second, third, etc.) of \(f(x)\).
5. Prove that if \(f\) is a polynomial function, then \(f'(x)\) is also a polynomial function.
(6) Prove that if \( f \) is a polynomial function of degree \( d \), then \( f^{(d+1)}(x) = 0 \) for all real numbers \( x \).

(7) Prove that if \( f \) is a polynomial function of degree \( d \), then \( f^{(d)} \) is a linear function.

(8) Let \( p \) be a polynomial function of degree \( d \) and let \( k \leq d \) be a nonnegative integer. What kind of function is \( f^{(k)} \)?

(9) Prove that if \( f(x) = x^{1/2} \) and \( a > 0 \), then \( f'(a) = \frac{1}{2\sqrt{a}} \).

(10) Prove that if \( f(x) = 1/x \) and \( a \neq 0 \), then \( f'(a) = -\frac{1}{a^2} \).

(11) Prove Theorem 3.4.

(12) Prove Theorem 3.5.

(13) Let \( f(x) = 3x^3 - 4x^2 + x - 2 + 4e^x \). Compute \( f'(x) \).

(14) Let \( f(x) = 1/\sqrt{x} \). Use the remark after the proof of Theorem 3.2 to compute \( f'(x) \).

(15) Let \( f(x) = x^2 - 2x + 7 \), and let \( g(x) = e^x \). Compute \((f + g)'(x)\) and \((f - g)'(x)\).

(16) Is there a function \( f \) that is not identically zero such that \( f^{(k)}(x) = f(x) \) for all \( x \)?

(17) Could it happen that \( f \) and \( g \) are two different functions, but \( f'(x) = g'(x) \) for all \( x \)?

(18) Could it happen that \( f \) and \( g \) are two different functions, \( f' \) and \( g' \) are two different functions, but \( f''(x) = g''(x) \) for all \( x \)?

(19) Could it happen that \( f \) and \( g \) are two polynomial functions of different degree, and \( f'(x) = g'(x) \) for all \( x \)?

(20) Is there a polynomial function \( f(x) \) that is not identically zero such that there is a real number \( x \) for which \( f^{(k)}(x) \) does not depend on \( k \)?

14. The Product and Quotient Rules

14.1. The Product Rule. We mentioned in the last section that, in general, \((fg)' \neq f'g'\). For instance, if \( f(x) = 2x + 1 \) and \( g(x) = x + 2 \), then \((fg)(x) = 2x^2 + 5x + 2\), so \((fg)'(x) = (2x^2 + 5x + 2)' = 4x + 5\), while \(f'(x) = 2\) and \(g'(x) = 1\), so \(f'(x)g'(x) = 2\).

It turns out that there is a rule to compute the derivative of a product; it is just a little bit more complicated than the limit law for products. This is the focus of our first theorem in this section.

**Theorem 3.6.** Let \( f \) and \( g \) be two functions that are differentiable at \( a \). Then \( fg \) is differentiable at \( a \), and

\[(fg)'(a) = f(a)g'(a) + f'(a)g(a).\]
Proof. By definition, we have

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

The crucial idea is to decompose the difference \( f(x)g(x) - f(a)g(a) \) as \( (f(x)g(x) - f(x)g(a)) + (f(x)g(a) - f(a)g(a)) \) in the numerator of the right-hand side of (3.2).

Using this idea, we obtain from Equation (3.2)

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

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

\[
= \lim_{x \to a} f(x) \cdot \frac{g(x) - g(a)}{x - a} + \lim_{x \to a} g(x) \cdot \frac{f(x) - f(a)}{x - a}
\]

\[
= f(a)g'(a) + g(a)f'(a).
\]

Example 3.2. The derivative of \( h(x) = x^2 e^x \) can be computed as follows. Let \( f(x) = x^2 \) and \( g(x) = e^x \). Then

\[
h'(x) = (fg)'(x)
\]

\[
= f(x)g'(x) + f'(x)g(x)
\]

\[
= x^2(e^x)' + (x^2)'e^x
\]

\[
= x^2e^x + 2xe^x
\]

\[
= e^x(x^2 + 2x).
\]

14.2. The Quotient Rule. The rule for the derivative of the quotient of two functions is a little bit more complicated than that for the derivative of the product of two functions. Though more complex, both the rule and its proof bear some similarity to the rule given in Theorem 3.6.

Theorem 3.7. Let \( f \) and \( g \) be two functions that are differentiable at \( a \) and let us assume that \( g(a) \neq 0 \). Then \( f/g \) is differentiable at \( a \), and we have

\[
\left( \frac{f}{g} \right)'(a) = \frac{g(a)f'(a) - f(a)g'(a)}{g(a)^2}.
\]

Proof. By definition, we have

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

(3.3)
Let us multiply both the numerator and the denominator of the right-hand side by \( g(x)g(a) \) to get
\[
\left( \frac{f}{g} \right)'(a) = \lim_{x \to a} \frac{f(x)g(a) - f(a)g(x)}{(x - a)g(x)g(a)}.
\]
Now transform the numerator of the right-hand side by subtracting and then adding \( g(a)f(a) \) to get
\[
\left( \frac{f}{g} \right)'(a) = \lim_{x \to a} \frac{f(x)g(a) - g(a)f(a) + g(a)f(a) - f(a)g(x)}{(x - a)g(x)g(a)}
\]
\[
= \lim_{x \to a} \frac{g(a)}{g(x)g(a)} \cdot \frac{f(x) - f(a)}{x - a} - \lim_{x \to a} \frac{f(a)}{g(x)g(a)} \cdot \frac{g(x) - g(a)}{x - a}
\]
\[
= \frac{g(a)f'(a) - f(a)g'(a)}{g(a)^2}.
\]

\( \square \)

Theorem 3.7 now enables us to compute the derivative of rational functions.

**Example 3.3.** Let \( h(x) = (x + 3)/(x^2 + 1) \). Find \( h'(x) \).

**Solution:** Let \( f(x) = x + 3 \) and let \( g(x) = x^2 + 1 \). Then \( f'(x) = 1 \) and \( g'(x) = 2x \). So, by Theorem 3.7, we have
\[
h'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2} = \frac{x^2 + 1 - (x + 3)2x}{x^2 + 2x^2 + 1} = \frac{-x^2 - 6x + 1}{x^4 + 2x^2 + 1}.
\]

\( \square \)

14.3. Exercises.

1. Let \( h(x) = e^x x^3 \). Find \( h'(x) \) and \( h''(x) \).
2. Let \( f(x) = (2x + 7)e^x \). Compute \( f'(x) \).
3. Find a rule to compute \( (f^2)'(x) \).
4. Find a rule to compute \( (f^3)'(x) \).
5. Find a rule to compute \( (1/f)'(x) \).
6. Use the result of the previous exercise to prove a formula for \( g'(x) \) if \( g(x) = x^n \) for a negative integer \( n \).
7. Let \( g(x) = e^{-x} \). Find \( g'(x) \).
8. Let \( h(x) = x/e^x \). Find \( h'(x) \).
9. Let \( f(x) = \frac{1 - 2x}{1 + 3x} \). Compute \( f'(x) \).
10. Let \( f(x) = \frac{2 - 3x}{1 - 5x} \). Compute \( f'(x) \).
11. Let \( f(x) = e^{2x}/(x + 2) \). Compute \( f'(x) \).
12. Let \( g(x) = e^{2x} \). Use the product rule to compute \( g'(x) \).
13. Let \( g(x) = e^{-2x} \). Compute \( g'(x) \). Try to find three different ways to obtain your result.
(14) Let \( h(x) = (e^x + 1)(e^x + 2) \). Compute \( h'(x) \).

(15) Let \( g(x) = (x - 3)/(e^x + 1) \). Compute \( g'(x) \).

(16) Let \( f(x) = (2x + 3)/(4x + 7) \). Compute \( f'(x) \). Try to find two different ways of getting the same answer.

(17) Let \( f(x) = 1/(1 - x) \). Find \( f'(x) \).

(18) Let \( f(x) = 1/(1 - x)^n \). Find \( f'(x) \).

(19) Let \( f(x) = g(x)h(x) \), where \( g \) is a polynomial function of \( x \), and \( h(x) = e^x \). Prove that \( f'(x) \) and \( f''(x) \) are each equal to the product of a polynomial function and the function \( h(x) = e^x \).

(20) Prove that if \( f(x) \) is a rational function, then \( f'(x) \) is also a rational function.

15. Derivatives of Trigonometric Functions

In this section, we show how to compute the derivatives of trigonometric functions. First, we compute \((\sin x)'\). This will be a somewhat lengthy procedure, due to the fact that this is the first trigonometric function we will differentiate and we will have to apply new methods. However, once we know the derivatives of \( \sin x \) and \( \cos x \), it will be much simpler to deduce the derivatives of other trigonometric functions, since those functions can be obtained from \( \sin \) and \( \cos \), and then the various differentiation rules can be used.

**Theorem 3.8.** We have \((\sin x)' = \cos x\).

**Proof.** Recall the identity \(\sin(a + b) = \sin a \cos b + \sin b \cos a\). We have

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

\[
= \lim_{h \to 0} \left( \sin x \frac{\cos h - 1}{h} + \sin h \frac{\cos x}{h} \right) = \sin x \lim_{h \to 0} \frac{\cos h - 1}{h} + \cos x \lim_{h \to 0} \frac{\sin h}{h}.
\]

Note that as, \( h \) approaches 0, we certainly have \( \lim_{h \to 0} \sin x = \sin x \) and \( \lim_{h \to 0} \cos x = \cos x \), since these functions do not even depend on \( h \).

There remains the task of computing the two nontrivial limits

\[
\lim_{h \to 0} \frac{\cos h - 1}{h} \quad \text{and} \quad \lim_{h \to 0} \frac{\sin h}{h}.
\]

We will carry out this task in two lemmas.
Lemma 3.1. We have
\[
\lim_{h \to 0} \frac{\sin h}{h} = 1.
\]

Proof. Let us consider a circle with unit radius and a regular \(n\)-gon whose center is at the center \(O\) of the circle and whose \(n\) vertices are all on the unit circle. Then the area of the circle is \(\pi\), and the area of the \(n\)-gon is \(n \cdot \frac{1}{2} \cdot \sin \alpha\), where \(\alpha = \frac{2\pi}{n}\) is the angle \(AOB\), with \(A\) and \(B\) being adjacent vertices of our \(n\)-gon.

Considering just \(1/n\) of both the circle and the \(n\)-gon, we see that the area of the triangle \(AOB\) is \((\sin \alpha)/2\), and the area of \(1/n\) of the circle bordered by the lines \(AO, BO\), and the arc \(AB\) is \(\pi \cdot \alpha/(2\pi) = \alpha/2\). So the ratio of the two areas is
\[
\frac{(\sin \alpha)/2}{\alpha/2} = \frac{\sin \alpha}{\alpha}.
\]

On the other hand, as \(n\) gets larger and larger, \(\alpha\) gets smaller and smaller, while the area of the \(n\)-gon gets closer and closer to the area of the circle. Hence, their ratio, \(\sin \alpha/\alpha\), will get arbitrarily close to 1 and stay arbitrarily close to 1.

\[\square\]

Lemma 3.2. The equality
\[
(3.4) \quad \lim_{h \to 0} \frac{\cos h - 1}{h} = 0
\]
holds.

Proof. We will manipulate the expression \((\cos h - 1)/h\) so that we can use the result of Lemma 3.1. First, we multiply both the numerator and the denominator by \(\cos h + 1\) to get
\[
\frac{\cos h - 1}{h} = \frac{\cos^2 h - 1}{h(1 + \cos h)} = -\frac{\sin^2 h}{h(1 + \cos h)}.
\]

Therefore, we have
\[
\lim_{h \to 0} \frac{\cos h - 1}{h} = -\lim_{h \to 0} \frac{\sin^2 h}{h(1 + \cos h)} = -\left(\lim_{h \to 0} \frac{\sin h}{h} \cdot \frac{\sin h}{1 + \cos h}\right) = -\left(\lim_{h \to 0} \frac{\sin h}{h}\right) \cdot \left(\lim_{h \to 0} \frac{\sin h}{1 + \cos h}\right) = (-1) \cdot 0 = 0.
\]

\[\square\]
We can now finish the proof of Theorem 3.8. At the end of the first displayed chain of equations in that proof, we saw that

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

The previous two lemmas showed that, on the right-hand side, the first limit is 0 and the second limit is 1, so \((\sin x)' = \cos x\) as claimed. \(\square\)

The following theorem can be proved by very similar methods.

**Theorem 3.9.** The equality \((\cos x)' = -\sin x\) holds.

You are asked to prove this theorem in Exercise 15.1.1.

Now that we have the derivatives of \(\sin\) and \(\cos\), the derivatives of other trigonometric functions can be obtained by simply using the quotient rule. The next theorem shows an example of this.

**Theorem 3.10.** We have \((\tan x)' = \sec^2 x\).

**Proof.** Note that \(\tan x = \sin x / \cos x\), so we can apply the quotient rule. This leads to

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

The derivatives of the other three trigonometric functions are given in the exercises.

**15.1. Exercises.**

1. Prove that \((\cos x)' = -\sin x\).
2. Prove that \((\cot x)' = -\csc^2 x\).
3. Prove that \((\csc x)' = -\csc x \cot x\).
4. Prove that \((\sec x)' = \sec x \tan x\).
5. Let \(h(x) = x \sin x\). Find \(h'(x)\).
6. Let \(h(x) = (x^2 - 2x + 3) \cos x\). Find \(h'(x)\).
7. Let \(h(x) = \frac{\tan x}{x+1}\). Find \(h'(x)\).
8. Let \(h(x) = e^x \cos x\). Find \(h'(x)\).
9. Let \(h(x) = e^x / \sin x\). Find \(h'(x)\).
(10) Let \( h(x) = e^{2x} \tan x \). Compute \( h'(x) \).

(11) Let \( h(x) = e^{-2x} \cot x \). Compute \( h'(x) \).

(12) Let \( h(x) = \frac{e^x + \sin x}{x^2 + 1} \). Find \( h'(x) \).

(13) Compute \( (\sin^2 x)' \).

(14) Compute \( (\cos^2 x)' \). Try to get the same answer in two different ways.

(15) Compute \( (\sin x \tan x)' \).

(16) Compute \( (\cot^2 x)' \).

(17) Compute \( (\tan^2 x)' \).

(18) Compute \( (\sec^2 x)' \). Try to get the same answer in two different ways.

(19) Compute \( (\sin^3 x)' \). You may want to use the result of exercise 13.

(20) Compute \( (\cos^3 x)' \). You may want to use the result of exercise 14.

16. The Chain Rule

16.1. The Derivative of the Composition of Two Functions. In previous sections, we learned how to compute the derivative of the sum, difference, product, and quotient of two functions. We still do not know how to compute the derivative of the composition of functions, such as \( h(x) = \sin(3x) \), \( t(x) = \sqrt{x^2 + 1} \), or \( r(x) = e^{\sin x} \). In this section, we will learn a rule, called the chain rule, that applies in these situations.

**Theorem 3.11 (Chain Rule).** Let \( h(x) = f(g(x)) \), where \( g \) is differentiable at \( x \) and \( f \) is differentiable at \( g(x) \). Then \( h \) is differentiable at \( x \), and we have

\[
h'(x) = f'(g(x))g'(x).
\]

In other words, we first differentiate the outside function at a point given by the inside function, then multiply the result by the derivative of the inside function.

The proof of the chain rule is somewhat technical, so we will postpone it until the end of this section. Now we will discuss some examples of the applications of the chain rule.

**Example 3.4.** Find the derivative of \( h(x) = \sin(3x) \).

**Solution:** Let \( f(x) = \sin x \) and let \( g(x) = 3x \). Then \( h(x) = f(g(x)) \), so, by the chain rule, we have

\[
h'(x) = f'(g(x)) \cdot g'(x) = (\cos(3x)) \cdot 3 = 3 \cos(3x).
\]
**Example 3.5.** Let \( h(x) = \sqrt{x^2 + 1} \). Find \( h'(x) \).

**Solution:** Recall that in Section 13, we mentioned that the identity \((x^n)' = nx^{n-1}\) holds for any nonzero real number. Therefore, selecting \( n = 1/2 \), we get that \((\sqrt{x})' = (x^{1/2})' = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}\).

Now we can prove the statement of the example. Let \( f(x) = \sqrt{x} \), and let \( g(x) = x^2 + 1 \). Then \( h(x) = f(g(x)) \), so, by the chain rule, we have

\[
h'(x) = f'(g(x))g'(x) = \frac{1}{2\sqrt{x^2 + 1}} \cdot 2x = \frac{x}{\sqrt{x^2 + 1}}.
\]

Sometimes the chain rule is written in the Leibniz notation, that is, as

\[
\frac{dh}{dx} = \frac{dh}{dg} \cdot \frac{dg}{dx}.
\]

**16.2. Two Applications of the Chain Rule.**

**16.2.1. A Simple Way of Obtaining \((\cos x)’\).** Recall that in the last section, it took considerable time and effort to prove that \((\sin x)’ = \cos x\). Finding \((\cos x)’\) with similar methods is just as time-consuming. On the other hand, the chain rule enables us to compute \((\cos x)’\) faster.

Recall that \( \cos x = \sin(x + \frac{\pi}{2}) \). So we can write \( \cos x \) as the composition of two functions, namely \( \cos x = f(g(x)) \), with \( f(x) = \sin x \) and \( g(x) = x + \frac{\pi}{2} \). So the chain rule applies, and we get

\[
(\cos x)' = f'(g(x)) \cdot g'(x) = \cos\left(x + \frac{\pi}{2}\right) \cdot 1 = \cos x \cos \frac{\pi}{2} - \sin x \sin \frac{\pi}{2} = 0 - \sin x = -\sin x.
\]

**16.2.2. The Derivatives of Exponential Functions.** Recall that we defined the number \( e \) such that the derivative of the exponential function \( f(x) = e^x \) was \( f(x) \) itself. Now the chain rule enables us to compute the derivatives of exponential functions with any base.

**Theorem 3.12.** Let \( a \) be a positive real number and let \( h(x) = a^x \). Then we have

\[
h'(x) = a^x \ln a.
\]
16. THE CHAIN RULE

PROOF. Note that
\[ h(x) = a^x = (e^{\ln a})^x = e^{x \ln a}. \]
So we have succeeded in writing \( h \) as the composition of two functions, namely \( h(x) = f(g(x)) \), where \( f(x) = e^x \) and \( g(x) = x \ln a \). Therefore, the chain rule applies, and we get
\[ h'(x) = f'(g(x)) \cdot g'(x) = e^{x \ln a} \cdot \ln a = a^x \ln a. \]

\[ \square \]

16.3. Proof of the Chain Rule. It is time that we proved the chain rule.

PROOF OF THEOREM 3.11. As \( g \) is differentiable at \( x \), we know that
\[ \lim_{r \to 0} \left( \frac{g(x + r) - g(x)}{r} - g'(x) \right) = 0. \]
Set
\[ t = \frac{g(x + r) - g(x)}{r} - g'(x). \]
Note that \( t \) depends on \( r \), and as \( r \) approaches 0, \( t \) approaches 0. Similarly, let \( y = g(x) \). As \( f \) is differentiable at \( y \), we have
\[ \lim_{s \to 0} \left( \frac{f(y + s) - f(y)}{s} - f'(y) \right) = 0. \]
Set
\[ u = \frac{f(y + s) - f(y)}{s} - f'(y). \]
Again, note that \( u \) depends on \( s \) and that \( u \) approaches 0 as \( s \) approaches 0.

Now we undertake a series of manipulations of the preceding two equations. Our goal is to express \( f(g(x))' = \lim_{r \to 0} \frac{f(g(x + r)) - f(g(x))}{r} \) in terms of \( f'(g(x)) \) and \( g'(x) \).

Rearranging the equation that defines the variable \( t \) that we just introduced, we get
\[ g(x + r) = g(x) + (g'(x) + t)r. \]
Similarly, rearranging the equation that defines the variable \( u \), we get
\[ f(y + s) = f(y) + (f'(y) + u)s. \]
Now apply the function \( f \) to both sides of (3.7) to get
\[ f(g(x + r)) = f(g(x) + (g'(x) + t)r). \]
Observe that (3.8) holds for all \(y\) and \(s\), so, in particular, it holds when \(y = g(x)\) and \(s = (g'(x) + t)r\). Making these substitutions in (3.8), Equation (3.9) yields

\[
(3.10) \quad f(g(x + r)) = f(g(x)) + (f'(g(x)) + u)(g'(x) + t)r.
\]

We can now express the quotient \((f(g(x + r)) - f(g(x))) / r\) from the equality of the left-hand side of (3.10) and the expression in (3.11) as

\[
\lim_{r \to 0} \frac{f(g(x + r)) - f(g(x))}{r} = \lim_{r \to 0} \frac{(f'(g(x)) + u)(g'(x) + t)r}{r} = (f'(g(x)) + u)(g'(x) + t).
\]

Finally, we are in a position to compute the derivative we were looking for as the limit of the left-hand side as \(r\) approaches 0. We get

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

since both \(t\) and \(u\) approach 0 as \(r\) approaches 0.

16.4. Exercises.

(1) Let \(h(x) = (x^2 + 1)^5\). Find \(h'(x)\).
(2) Let \(h(x) = \cot(2x)\). Find \(h'(x)\).
(3) Let \(h(x) = \cos(2x + 8)\). Find \(h'(x)\).
(4) Let \(h(x) = \sin(x^2)\). Find \(h'(x)\).
(5) Let \(h(x) = \sin^3 x\). Find \(h'(x)\). Compare your result to the result of exercise 19 of the previous section. Which other exercises of the previous section can be solved by the chain rule?
(6) Let \(f(x) = 2^x + 3^x\). Find \(f'(x)\).
(7) Let \(h(x) = e^{\sin x}\). Find \(h'(x)\).
(8) Let \(h(x) = 2^{\cos x}\). Find \(h'(x)\).
(9) Let \(h(x) = e^{x^2 \sin x}\). Find \(h'(x)\).
(10) Let \(h(x) = \sqrt{x^2 + 2x + 7}\). Find \(h'(x)\).
(11) Let \(h(x) = \sin(e^x + 5x + 6)\). Find \(h'(x)\).
(12) Let \(h(x) = e^{\sqrt{x+1}}\). Find \(h'(x)\).
(13) Let \(h(x) = e^{\sin(x^2)}\). Find \(h'(x)\).
(14) Let \(h(x) = \sin(2x)\). Find \(h'(x)\). How could you get the same result without using the chain rule?
(15) Let \( h(x) = \cos(2x) \). Find \( h'(x) \). How could you get the same result without using the chain rule?

(16) Let \( h(x) = 2x^2 \). Find \( h'(x) \).

(17) Let \( h(x) = 1/(1-x) \). Find \( h'(x) \). Try to get the same answer in two different ways.

(18) Let \( h(x) = (2 + \tan x)^3 \). Find \( h'(x) \).

(19) Let \( h(x) = \sqrt{1-x^2} \). Find \( h'(x) \).

(20) Let \( h(x) = \sqrt{x + \sqrt{x + 1}} \). Find \( h'(x) \).

17. Implicit Differentiation

In the last several sections, we computed the derivatives of many different functions. Although these functions were different, they had one important feature in common. They were *explicitly* given. That is, they were given by a rule that directly described how \( f(x) = y \) is obtained from \( x \).

17.1. Tangent Lines to Implicitly Defined Curves. Sometimes we have to deal with curves that are given by a different kind of rule. Consider the curve given by the equation

\[
3.12 \quad x^3 + y^3 = 4xy.
\]

Let us say that we want to compute the slope of the tangent line to this curve at the point \((2, 2)\). If we could express \( y \) as a function of \( x \), we could simply take the derivative of that function at \( x = 2 \). However, it is not clear how to write \( y \) *explicitly* in terms of \( x \), even if \( 3.12 \) *implicitly* describes this dependence.

It is in these situations that we resort to *implicit* differentiation. Keep in mind that we do not need to explicitly know how \( y \) depends on \( x \), that is, we do not need an explicit expression for the function \( y(x) \); we only need to know the derivative \( dy/dx \) of that function at \( x = 2 \).

Consider Equation \( 3.12 \), and differentiate both sides with respect to the variable \( x \) to get

\[
\frac{d}{dx} \left( x^3 + y^3 \right) = \frac{d}{dx} (4xy).
\]

Now recall that \( y = y(x) \) is a function of \( x \). So, when computing \( (d/dx)y^3 \) on the left-hand side, we need to use the chain rule. On the right-hand side, we need to use the product rule and the chain rule. Using these rules, we get

\[
3x^2 + 3y^2 \frac{dy}{dx} = 4y + 4x \frac{dy}{dx}.
\]
Expressing \( dy/dx \) from this equation, we get

\[
\frac{dy}{dx} = \frac{4y - 3x^2}{3y^2 - 4x}.
\]

At the point \((2, 2)\), the right-hand side is \(-4/4 = -1\), so the slope of the tangent line at \((2, 2)\) is \(-1\).

Note that the fact that the tangent line at \((2, 2)\) has slope \(-1\) makes (intuitively) perfect sense, since the curve in question is symmetric in \(x\) and \(y\). That is, if \((x, y)\) is on the curve, then \((y, x)\) is also on the curve.

17.2. Derivatives of Inverse Trigonometric Functions. One place where implicit differentiation is a very powerful tool is in the computation of the derivatives of inverse trigonometric functions. Recall that \(\tan^{-1} x = y\) is the function that is the inverse of the restriction of the function \(\tan x\) to the interval \((-\pi/2, \pi/2)\). That is, if

\[
\tan^{-1} x = y,
\]

then

\[
x = \tan y,
\]

where \(y \in (-\pi/2, \pi/2)\).

Our goal is to determine

\[
\frac{d}{dx} \tan^{-1} x = \frac{dy}{dx}.
\]

To that end, let us take the derivative of both sides of (3.13) with respect to \(x\). Recalling that

\[
\frac{d}{dz} \tan z = \sec^2 z \quad \text{and} \quad y = y(x),
\]

we get

\[
1 = \sec^2 y \cdot \frac{dy}{dx}.
\]

Solving for \(dy/dx\) and recalling the identity \(\sec^2 z = 1 + \tan^2 z\), we obtain

\[
\frac{dy}{dx} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}.
\]

In other words, we proved the surprisingly simple formula

\[
(\tan^{-1} x)' = \frac{1}{1 + x^2}.
\]
This formula is interesting for two reasons. First, it is surprisingly simple. Second, it does not even contain trigonometric functions. Imagine trying to get this result without implicit differentiation, using just the definition of derivatives.

You will be asked to compute the derivatives of the other inverse trigonometric functions in the exercises.

17.3. Exercises.

(1) Let $C$ be the circle given by the equation $x^2 + y^2 = 169$. Use implicit differentiation to find the slope of the tangent line to $C$ at the point $(5, 12)$.

(2) Let $E$ be the ellipse given by the equation $x^2 + xy + y^2 = 108$. Use implicit differentiation to find the slope of the tangent line to $E$ at the point $(6, 6)$.

(3) Let $y$ be implicitly defined by the equation $x^5 + y^5 = 1$. Compute $dy/dx$.

(4) Let $y$ be implicitly defined by the equation $y = \sin(xy)$. Compute $dy/dx$.

(5) Let $y$ be implicitly defined by the equation $\sqrt{x} + \sqrt{y} = y$. Compute $dy/dx$.

(6) Let $y$ be implicitly defined by $\sqrt{x} + \sqrt{y} = 4$. Compute $dy/dx$.

(7) Let us assume that $h(x) + x \cos h(x) = x^3$ for all real numbers $x$. Is this sufficient information to determine the value of $h'(0)$?

(8) Let us assume that $h(x) + e^{h(x)} = 5$ for all real numbers $x$. Prove that then $h$ must be a constant function.

(9) Prove that $(\sin^{-1} x)' = \frac{1}{\sqrt{1-x^2}}$.

(10) Prove that $(\cos^{-1} x)' = -\frac{1}{\sqrt{1-x^2}}$.

(11) Prove that $(\cot^{-1} x)' = -\frac{1}{1+x^2}$.

(12) Prove that $(\sec^{-1} x)' = \frac{1}{x\sqrt{x^2-1}}$.

(13) Prove that $(\csc^{-1} x)' = -\frac{1}{x\sqrt{x^2-1}}$.

(14) Compute $(\tan^{-1}\sqrt{x})'$.

(15) Compute $(\cos^{-1}(2x + 0.1))'$.

(16) Compute $(\sin^{-1}(x^2))'$.

(17) Compute $(\sin^{-1}(1/x))'$.

(18) Compute $(\tan^{-1}(2x))'$.

(19) Compute $(\csc^{-1}(x/3))'$.

(20) Compute $(\sec^{-1}(x - 0.01))'$.
18. Derivatives of Logarithmic Functions

18.1. The Formula for \((\log_a x)\)' . As another powerful application of implicit differentiation, we compute the derivative of the function \(f(x) = \ln x\).

**Theorem 3.13.** We have
\[
(\ln x)' = \frac{1}{x}.
\]

**Proof.** Set \(y = \ln x\). Then \(e^y = x\). Differentiating both sides with respect to \(x\), we get
\[
e^y \cdot \frac{dy}{dx} = 1,
\]
\[
\frac{dy}{dx} = \frac{1}{e^y}.
\]
However, \(e^y = x\) by definition, so
\[
\frac{dy}{dx} = \frac{1}{x}
\]
as claimed.

Note that the function \(\ln x\) is defined for positive values of \(x\). If \(x < 0\), then the function \(y = \ln(-x)\) is defined. The reader is invited to practice the method of implicit differentiation by showing that \((\ln(-x))' = 1/x\) for all negative real numbers \(x\). In other words, if \(y = \ln(|x|)\), then \(dy/dx = 1/x\) as long as \(x\) is a nonzero real number.

It is now a breeze to determine the derivative of logarithmic functions of any base.

**Corollary 3.1.** Let \(a \neq 1\) be a fixed positive real number. Then
\[
(\log_a x)' = \frac{1}{x \ln a}.
\]

**Proof.** Note that
\[
x = (e^{\ln a})^{\log_a x} = e^{(\ln a)(\log_a x)}.
\]
So \(\ln x = (\ln a)(\log_a x)\) and
\[
f(x) = \log_a x = \frac{\ln x}{\ln a}.
\]
As \(\ln a\) is a constant, it follows that
\[
f'(x) = \frac{1}{\ln a} (\ln x)' = \frac{1}{x \ln a}
\]
as claimed.
18.2. The Chain Rule and In $x$. An interesting consequence of Theorem 3.13 is the following.

Corollary 3.2. Let $f(x)$ be a differentiable function that takes positive values only. Then

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

Proof. By the chain rule,

$$\frac{d}{dx} \ln f = \frac{d}{df} \cdot \frac{df}{dx} = \frac{f'(x)}{f(x)}.$$

Example 3.6. Let $f(x) = \cos x$. Then

$$\frac{d}{dx} \ln(\cos x) = -\frac{\sin x}{\cos x} = -\tan x.$$

18.3. Logarithmic Differentiation. Sometimes we need to compute the derivative of a complicated product. This is sometimes easier by taking the logarithm of the product, which will be a sum, and using implicit differentiation. This procedure, which is called logarithmic differentiation, has the inherent advantage that it deals with sums instead of products, and sums are much easier to differentiate than products.

Example 3.7. Let

$$y = \frac{x^3 \sqrt{x+1}}{\sqrt{x-2}}.$$

Compute $dy/dx$.

Solution: Taking logarithms, we get

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

Now taking derivatives with respect to $x$ and using Corollary 3.2, we have

$$\frac{dy}{dx} \cdot \frac{1}{y} = \frac{3}{x} + \frac{1}{2(x+1)} - \frac{1}{2(x-2)}.$$

Finally, we can solve this equation for $dy/dx$ to get

$$\frac{dy}{dx} = y \left( \frac{3}{x} + \frac{1}{2(x+1)} - \frac{1}{2(x-2)} \right) = \frac{x^3 \sqrt{x+1}}{\sqrt{x-2}} \cdot \left( \frac{3}{x} + \frac{1}{2(x+1)} - \frac{1}{2(x-2)} \right).$$
18.4. Power Functions Revisited. Recall that in an earlier section, we proved that if \( n \) is a fixed positive integer, then \( (x^n)' = nx^{n-1} \). We stated that this was the case for all nonzero real numbers \( n \), not just positive integers, but we have not proved that claim. Now we have the tools, namely logarithmic differentiation, to prove it.

**Theorem 3.14.** Let \( n \) be any nonzero real number. Then we have

\[
\frac{d}{dx} x^n = nx^{n-1}.
\]

**Proof.** Set \( y = x^n \). Let us assume for the case of simplicity that \( x \) is positive. Taking logarithms, we have

\[\ln y = n \ln x.\]

Differentiating both sides with respect to \( x \), we get

\[
\frac{dy}{dx} \cdot \frac{1}{y} = \frac{n}{x},
\]

Solving for \( dy/dx \) yields

\[
\frac{dy}{dx} = \frac{ny}{x} = \frac{nx^n}{x} = nx^{n-1}
\]

as claimed. \(\square\)

18.5. The Number \( e \) Revisited. Recall that we have defined the number \( e \), the base of the natural logarithm, as the number for which \( \lim_{h \to 0} (e^h - 1)/h = 1 \). Our new knowledge lets us express \( e \) more directly, as a limit.

Note that if \( f(x) = \ln x \), then \( f'(x) = 1/x \), so \( f'(1) = 1 \). By the definition of derivatives, this means that

\[
\lim_{h \to 0} \frac{\ln(1 + h) - \ln 1}{h} = 1.
\]

Observing that \( \ln 1 = 0 \) and using the power rule of logarithms, we get

\[
\lim_{h \to 0} \ln(1 + h)^{1/h} = 1,
\]

or, applying the exponential function \( e^z \) to both sides, we have

\[
\lim_{h \to 0} (1 + h)^{1/h} = e.
\]

Equivalently, setting \( x = 1/h \), we get

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

Either of the last two formulas can help to determine the approximate value 2.712828 of \( e \).
18.6. Exercises.

(1) Let \( h(x) = 2^x + 3^x \). Compute \( h'(x) \).

(2) Let \( h(x) = 4^{-x} \). Compute \( h'(x) \).

(3) Let \( h(x) = 2\sqrt{x} \). Compute \( h'(x) \).

(4) Let \( h(x) = e^{2x} \). Compute \( h'(x) \).

(5) Compute \( \frac{d}{dx} \ln(\sqrt{x + 1}) \).

(6) Compute \((\ln|x|)'\).

(7) Compute \((x^x)'\).

(8) Let \( h(x) = \sqrt[3]{x+1} \). Compute \( h'(x) \).

(9) Let \( h(x) = \sqrt[4]{x^2 + 2} \). Compute \( h'(x) \).

(10) Let \( h(x) = \sqrt[3]{2x + 3} \). Compute \( h'(x) \).

(11) Compute \( f'(x) \) if \( f(x) = x^4 \frac{3x^2 + 4}{x+1} \).

(12) Compute \( \lim_{x\to\infty} (1 + \frac{1}{x})^{2x} \).

(13) Compute \( \lim_{x\to\infty} (1 + \frac{1}{x})^{-x} \).

(14) Compute \( \lim_{x\to\infty} \left(1 - \frac{1}{x}\right)^x \).

(15) Let \( h(x) = \ln(\sin^{-1}x) \). Compute \( h'(x) \).

(16) Let \( h(x) = \sin^{-1}(\ln x) \). Compute \( h'(x) \).

(17) Let \( h(x) = \ln \left(\frac{1}{x^2}\right) \). Compute \( h'(x) \).

(18) Use logarithmic differentiation to find \( y'(x) = \frac{dy}{dx} \) if \( y(x) = x^{\ln{x}} \).

(19) Use logarithmic differentiation to find \( y'(x) = \frac{dy}{dx} \) if \( y(x) = x^{\sqrt{x}} \).

(20) Use logarithmic differentiation to find \( y'(x) = \frac{dy}{dx} \) if \( y(x) = x^{2\cos{x}} \).

19. Applications of Rates of Change

In this section, we consider a few applications of derivatives in various disciplines.

19.1. Physics. Recall that if an object moves along a line and the distance it covers in time \( t \) is described by the function \( s(t) \), then

\[
\frac{ds}{dt} = s'(t) = \lim_{h \to 0} \frac{s(t+h) - s(t)}{h}
\]

is the instantaneous velocity of the object at time \( t \).

We can take this concept one step further. If the object moves at a changing velocity, then the rate of change of the velocity itself can be important information. For instance, when considering a vehicle’s performance, we may be interested in how fast it can reach its top speed, not only what its top speed is.
The corresponding notion in physics is called \textit{acceleration}, and is denoted by $a(t)$. That is, keeping the previous notation, we have

\begin{equation}
    a(t) = v'(t) = \frac{dv}{dt} = s''(t).
\end{equation}

\textbf{Example 3.8.} The position of a particle is described by the equation

\begin{equation}
    s(t) = \frac{1}{3}t^3 - 3t^2 + 5t.
\end{equation}

Here \(s\) is measured in meters and \(t\) in seconds.

(I) What is the velocity of the particle after 3 seconds?

(II) Find the acceleration of the particle after 10 seconds.

(III) When does the particle move backward?

Additional questions about the movement of this particle will be given in the exercises.

\textbf{Solution:}

(I) The velocity of the particle is described by the function $v(t) = s'(t) = t^2 - 6t + 5$. This yields $v(3) = 9 - 18 + 5 = -4$. So the velocity of the particle after 3 seconds is \(-4\) m/s, meaning that the particle is moving backward at a speed of 4 meters per second after 3 seconds.

(II) The acceleration of the particle is given by the formula $a(t) = v'(t) = 2t - 6$. So, after 10 seconds, the particle is accelerating at 14 m/s\(^2\).

(III) The particle is moving backward when its velocity $v(t)$ is negative. That happens when $v(t) = t^2 - 6t + 5 = (t - 1)(t - 5) < 0$, that is, when $t \in (1, 5)$. In other words, the particle is moving backward between the first and fifth seconds.

\textbf{19.2. Economics.} Let us say that a company estimates that it costs \(C(x)\) dollars to produce \(x\) units of a new product. It is often the case that \(C(x)\), which is called the \textit{cost function}, can be described by a polynomial function, such as

\[C(x) = a + bx + cx^2 + dx^3.\]

The reason for this is as follows. There will be some costs, such as designing the product and obtaining permits, that will be present regardless of the number of units produced. These will be represented by the constant term \(a\). Then there will be costs, such as renting a location and buying supplies, that will be more or less in direct proportion to the number of units produced. These will be represented by the linear
term $bx$. Then there will be other factors, such as hiring workers, marketing the product, and organizing production, that will be in direct proportion to a higher power of $x$ as the differences in size turn into differences in kind. Taxes may factor in at an even higher rate.

Because the cost function $C(x)$ is not a linear polynomial, producing the 1001st unit does not cost of the same as producing the first unit or the 5001st unit. The cost of increasing production from $n$ units to $n + 1$ units, in other words, the cost of producing the $(n + 1)$th unit, can be computed by the formula

$$M(n) = C(n + 1) - C(n).$$

The marginal cost function $C'(x)$ describes how the cost function changes. In that, $C'(x)$ and $M(n)$ are similar. There is one important difference. As we know, the derivative $C'(x)$ is given by

$$\lim_{\Delta x \to 0} \frac{C(x + \Delta x) - C(x)}{\Delta x}.$$ 

However, it could well be that the smallest meaningful positive value of $\Delta x$ is 1, in case the products are such that fractional units do not make sense (e.g., automobiles). In that case, $\Delta x \to 0$ is impossible in its precise mathematical meaning; the closest that $\Delta x$ can get to 0 is when $\Delta x = 1$. In that case, however, the expression after the limit symbol in (3.17) simplifies to $C(x + 1) - C(x)$, justifying the approximation

$$M(x) = C(x + 1) - C(x) \approx C'(x).$$

**Example 3.9.** The cost function of a bottle of a new medication is given by $C(x) = 10^6 + 20x + 0.001x^2 + 0.000001x^3$. Find the approximate cost of producing the 101st and the 1001st bottles.

**Solution:** By the preceding discussion, we need to compute the function $C'(x)$. By the rules of differentiating a polynomial function, we get $C'(x) = 0.000003x^2 + 0.002x + 20$. So the 101st bottle costs $0.0003 \cdot 100^2 + 0.002 \cdot 100 + 20 = 20.23$ dollars to produce, while the 1001st bottle costs $0.000003 \cdot 1000^2 + 0.002 \cdot 100^2 + 20 = 43$ dollars to produce.

It is important to note that the result of the previous example, that is, the fact that it costs more to produce the 1001st bottle than the 101st bottle does not mean that the more bottles are produced, the more expensive it is to produce the average bottle. This is because the cost of producing the first bottle is astronomical, since $C(1) > 10^6$. Compared to that, the cost of each of the first thousand, or even, first
ten thousand bottles is very small, so the production of each of them will bring the cost of producing the average bottle down. (The cost of producing the average bottle if \( n \) bottles are produced is of course \( C(n)/n \).

In the exercises, you are asked to compare these results to the results obtained by using the formula \( C(n + 1) - C(n) \).

19.3. Exercises.

(1) Consider the particle of Example 3.8. After 6 seconds, how far from its starting point is that particle? In what direction?

(2) Consider the particle of the previous exercise. Are there any moments when the particle is not moving?

(3) The location of an object moving vertically is described by the function \( s(t) = t - \frac{t^2}{5} \) for \( t \in [0, 5] \), where time is measured in seconds and distance is measured in meters. When will the object have an instantaneous velocity of 0.2 m/s?

(4) Consider the object of the previous exercise. When does it have the greatest speed going up? When does it have the greatest speed going down?

(5) Consider the object of Exercise 19.3.3. Will its acceleration ever be 1 m/s²?

(6) Consider the object of Exercise 19.3.3. When will its acceleration be negative?

(7) Use the formula \( M(n) = C(n + 1) - C(n) \) to find the cost of producing the 101st and 1001st units in Example 3.9. Compare your results with the estimates that we found using the function \( C'(x) \).

(8) Two race cars speed up from a standing start to 60 m/s so that each car has constant acceleration. The first car reaches one-third of its top velocity in 4 seconds, while the second car reaches one-fourth of its top velocity in 3 seconds. Which car will have covered more distance by the time it reaches its top velocity?

(9) A ball is rolling down a slope so that its distance from its starting point is described by the function \( s(t) = 2t^2 + 6t \), where \( 0 \leq t \leq 10 \), the time \( t \) is measured in seconds, and the distance \( s(t) \) is measured in meters. What will be the velocity of the ball after 3 seconds?

(10) Consider the ball of the previous exercise. When will its velocity reach 40 m/s?
(11) Consider the ball of the previous exercise and describe its acceleration as a function of the time $t$ passed from the start of the ball’s movement.

(12) The cost function for a company to produce $x$ bicycles is $C(x) = 2400 + 3x + 0.6x^2 + 0.002x^3$. Find the marginal cost function of this product.

(13) Consider the cost function $C(x)$ of the previous exercise and use it to determine the actual cost of producing the 1001st bicycle.

(14) The cost function for a company to produce $x$ laptops is $C(x) = 1500 + 2x + 0.4x^2 + 0.01x^3$. Find the marginal cost function for this product.

(15) Consider the cost function of the previous exercise. Explain the meaning of $C'(200)$.

(16) Let us say that the function $f(t)$ describes the growth of a certain bacteria population over time. That is, $f(0)$ is the size of the population at the beginning of the observation period, while $f(t)$ is the size of the population $t$ hours after that. Explain why $f'(t)$ describes the growth rate of the population in the moment of time corresponding to $t$.

(17) A certain insect population has been exposed to an insecticide, which results in the population changing according to the function $f(t) = 10,000 - 1000t - 500t^2$, where $t$ is measured in hours. Find the growth rate of the insect population after 1 hour and after 5 hours.

(18) A certain bacteria population has an initial size of 1000, and it doubles in each hour for the next 10 hours. Describe the growth of this bacteria population by a function, then use that function to determine the growth rate of the population after 3.5 hours.

(19) A certain rumor spreads according to the function

\[ P(t) = \frac{1}{1 + 30e^{-2t}}. \]

Here $P(t)$ is the proportion of the observed population that heard the rumor $t$ days after it started circulating. Explain why $P'(t)$ should be defined as the rate at which the rumor is spreading. Then compute that rate at $t = 4$.

(20) Consider the rumor discussed in the previous exercise. If (3.19) remains correct as $t$ goes to infinity, what fraction of the observed population will eventually hear the rumor?
20. Related Rates

20.1. Preliminaries. An intuitive idea of the notion of related rates comes from a simple fact of everyday life: If there are two related quantities that are changing with time, then their rates of change should also be related. For example, the volume $V$ of water in a pool of area $20 \text{ m}^2$ is related to the water level $h$ (the pool depth in meters) as $V = 20h$. Suppose the water level is low and needs to be increased. A hose is put into the pool that can pump water at a rate of $0.2 \text{ m}^3/\text{h}$. At what rate does the water level increase? The volume and the water level are both functions of time, $V = V(t)$ and $h = h(t)$. For every instance of time $t$, their values are related as $V(t) = 20h(t)$ and so must be their derivatives or rates of change:

$$(3.20) \quad V(t) = 20h(t) \implies V'(t) = 20h'(t).$$

Now the question is easy to answer. Since $V'(t) = 0.2 \text{ m}^3/\text{h}$, $h'(t) = V'(t)/20 = 0.01 \text{ m/h} = 1 \text{ cm/h}$. The water level rises by 1 cm every hour. A somewhat practical estimate! You would know exactly when to come back and turn off the water if you needed an inch or so of the water level increase. Apparently, the same idea of related rates would work for lowering the water level after rain.

20.2. Units. It is important to bring all the quantities to the same system of units. For example, in the above problem the pool area is often given in square feet, for example, $200 \text{ ft}^2$, while the pumping rate is given in gallons per hour, for example, $V' = 60 \text{ gal/h}$. One gallon is $3.785 \cdot 10^{-3} \text{ m}^3$ and therefore $V' = 60 \cdot 3.785 \cdot 10^{-3} = 0.2271 \text{ m}^3/\text{h}$. One square foot is $9.29 \cdot 10^{-2} \text{ m}^2$, so the pool area is $200 \cdot 9.29 \cdot 10^{-2} = 18.58 \text{ m}^2$. Hence, $h' = 0.2271/18.58 \approx 1.2 \text{ cm/h}$. In 1999, NASA lost a $125$ million Mars orbiter because a Lockheed Martin engineering team used English units of measurement while the agency’s team used the more conventional metric system for a key spacecraft operation.

20.3. Formal Definition of Related Rates.

**Definition 3.2 (Related quantities).** Two quantities $y$ and $x$ are said to be related if there is a function $f$ such that $y = f(x)$.

In the previous example, $V = f(h) = 20h$. Suppose now that the quantities $y$ and $x$ are functions of another variable $t$ (e.g., $t$ is time): $x = x(t)$ and $y = y(t)$. Then the rate of change of $x$ or $y$ with respect to $t$ is nothing but the derivative $x'(t)$ or $y'(t)$. The problem of “related
rates” can now be cast in the proper mathematical terms: What is the relationship between the derivatives \(x'(t)\) and \(y'(t)\) if the values of \(x(t)\) and \(y(t)\) are related by \(y = f(x)\)? The values of the functions \(x(t)\) and \(y(t)\) are related as \(y(t) = f(x(t))\) for any \(t\). Taking the derivative of both sides with respect to \(t\) by means of the chain rule (Theorem 3.11), we obtain a generalization of (3.20):

\[
y(t) = f(x(t)) \implies y'(t) = f'(x(t))x'(t).
\]

Equation (3.21) establishes the sought-after relation between the rates \(y'\) and \(x'\). However, it seems somewhat different from (3.20): The rates are still proportional to one another, but the proportionality coefficient \(f'(x)\) is no longer a constant, but a function. How do we use it? Take a particular value of \(t = t_0\). Let the values of \(x\) and \(y\) at \(t = t_0\) be \(x_0 = x(t_0)\) and \(y_0 = y(t_0)\). The number \(a = f'(x_0)\) can be calculated. Then the equality \(y'(t_0) = ax'(t_0)\) determines the relation between the rates \(y'\) and \(x'\) at the instance when \(x\) has the value \(x_0\) (or \(y\) has the value \(y_0 = f(x_0)\)).

**Example 3.10.** Let a laser pointer be positioned at a distance \(D = 1\ m\) from a wall. The pointer can be rotated so that the bright spot created by the laser beam travels horizontally on the wall.

(I) At what speed does the bright spot travel along the wall if the pointer revolves at a constant rate \(\omega\ \text{rad/s}\)?

(II) At what direction of the laser beam does the bright spot travel at the speed \(v = 4\pi\ \text{m/s}\) if \(\omega = \pi\ \text{rad/s}\)?

**Solution:**

(i) The analysis of any problem on related rates must begin with defining the quantities whose rates are being studied. In other words, one has to answer the question: How are these quantities measured? The orientation of the laser beam can be described by the angle \(\varphi\) between the perpendicular to the wall and the laser beam. The position of the bright spot may be set by the distance \(y\) traveled by it from the point on the wall when \(\varphi = 0\), that is, when the laser beam is perpendicular to the wall. If the pointer rotates, the angle becomes a function of time, \(\varphi = \varphi(t)\), and so does the position of the bright spot, \(y = y(t)\). Thus, the question is about the relation between the rates \(y'(t) = v\) (the speed at which the bright spot travels) and \(\varphi'(t) = \omega\) (the rate at which the pointer rotates).

(ii) The next step is to find a function that determines the relation between the quantities of interest, that is, between the distance
Figure 3.1. A laser pointer is positioned at a distance $D$ from a wall and rotates clockwise. Its beam makes a bright spot that moves with the speed $v$ to the right along the wall. The laser beam direction is determined by the angle $\varphi$ and the position of the bright spot is determined by the distance $y$.

$y$ and the angle $\varphi$: $y = f(\varphi)$. It is clear that $D$ and $y$ are related as the catheti of the right triangle whose hypotenuse is the laser beam: $y = D \tan \varphi = f(\varphi)$.

(iii) Once the relation between the quantities of interest has been established, the relation between their rates can be found. Since $(\tan \varphi)' = 1/\cos^2 \varphi$, Equation (3.21) yields

$$y = D \tan \varphi \quad \Rightarrow \quad y' = \frac{D}{\cos^2 \varphi} \varphi' \quad \Rightarrow \quad v = \frac{D}{\cos^2 \varphi} \omega.$$  

The first question is answered.

(iv) Note that the rate $y' = v$ is not constant even if the rate $\varphi' = \omega$ is constant. To answer the second question, one has to find the value of $\varphi$ when $v = 4\pi$ m/s, $D = 1$ m, and $\omega = \pi$ rad/s. It follows from Equation (3.22) that

$$\cos^2 \varphi = \frac{D \omega}{v} = \frac{1}{4} \quad \Rightarrow \quad \varphi = \frac{\pi}{3}.$$
that is, the bright spot moves at the speed $4\pi \text{ m/s}$ when the laser beam makes $60^\circ$ with the perpendicular to the wall. 

20.4. Can Anything Travel Faster Than Light? The solution (3.22) has an interesting feature. When $\varphi$ approaches $90^\circ$, that is, the laser beam is getting closer to being parallel to the wall, the cosine, $\cos \varphi$, tends to 0 in Equation (3.22), and hence the rate $y' = v$ grows unboundedly. It seems like just with merely a laser pointer, a superluminal object can be created in a lecture hall! Let us investigate this. The speed of light is $c \approx 300,000 \text{ km/s} \approx 186,000 \text{ mi/sec}$. The light can make a trip around the world in merely 0.13 seconds! Example 3.10 is now supplemented by two additional questions:

(III) Is it possible that $v$ can exceed the speed of light? If so, at which direction of the laser beam does it happen?

(IV) At which position of the bright spot does it happen?

The answers read:

(III) Setting $D = 1 \text{ m} = 10^{-3} \text{ km}$ (watch the units: all distances are now in kilometers!) and $v = c = 3 \cdot 10^5 \text{ km/s}$, the angle at which the bright spot exceeds the speed of light satisfies the equation $\cos^2 \varphi = D\omega/c \approx 1.05 \cdot 10^{-8}$, and hence $\varphi \approx 89.99414^\circ$. So the bright spot becomes superluminal if $\varphi > 89.99414^\circ$!

(IV) Since $y = D \tan \varphi$, $v > c$ if $y > 9772 \text{ m}$. Well, a lecture hall appears to be a “bit” small for this experiment! Take a Dremel miniature grinder (sold in Lowe’s stores) for which $\omega \approx \pi \cdot 10^3 \text{ rad/s}$ (it can be used to rotate the pointer), and set $D = 0.1 \text{ m}$, then $v > c$ if $y > 98 \text{ m}$; not yet exactly a lecture hall experiment, but it can be managed on the campus!

Einstein’s theory of relativity states that no material object can travel faster than light. Has a counterexample to Einstein’s theory just been found? The answer is “no.” In the motion of the bright spot, no material object actually moves along the wall. Bright spots at $y$ and $y + \Delta y$ are created by different portions of the laser beam that are emitted by the laser at two distinct moments of time. A lump of light that arrived at $y$ was reflected by the wall (that is why we see the bright spot!), and hence it could not appear at the next position $y + \Delta y$ (at this position arrived a different lump of light emitted by the laser at a later time). So the rate $\Delta y/\Delta t$ cannot possibly be associated with the motion of any material object along the wall.
20.5. Related Problem. The next time you watch a Florida sunset, look at your shadow. Does there exist a position of the Sun above the horizon at which your shadow extends faster than the speed of light?

20.6. More Than Two Related Rates. There are situations when several quantities are related among themselves. If these quantities become functions of a variable \( t \), then their rates are linearly related. A proof of this statement is given in Calculus 3, where functions of several variables are studied. However, the basic idea of finding relations between the rates has not changed: They are obtained by differentiating the relations between the quantities in question with respect to \( t \). The procedure is illustrated in the following example.

**Example 3.11.** Consider a rectangle with sides \( x \) and \( y \). Suppose that \( x \) and \( y \) change with time. Find their rates of change when \( x = 3 \) cm and \( y = 1 \) cm if, at that moment, the area of the rectangle decreases at a rate of \( 2 \) cm\(^2\)/s while the perimeter does not change.

**Solution:**

(i) There are four quantities involved: the rectangle dimensions \( x \) and \( y \), the area \( S \), and the perimeter \( P \).

(ii) There are two relations between them:

\[
S = xy, \quad P = 2(x + y).
\]

(iii) If \( x = x(t) \) and \( y = y(t) \), then \( S(t) = x(t)y(t) \) and \( P(t) = 2(x(t) + y(t)) \). Using the derivative of the product and the sum of two functions, the linear relations between the rates are obtained

\[
S' = x'y + xy', \quad P' = 2(x' + y').
\]

(iv) Since \( P' = 0 \) (the perimeter does not change), \( x' = -y' \) and \( S' = (x - y)y' \). Now let \( S' = -2 \) cm\(^2\)/s because \( S \) decreases (\( S' \) must be negative). With \( x = 3 \) cm and \( y = 1 \) cm, one has \(-2 = (3 - 1)y' \) and \( y' = -1 \) cm/s. It then follows that \( x' = -y' = 1 \) cm/s. \( \Box \)

20.7. Exercises.

1. Consider a triangle with vertices \( A, B, C \) such that \(|AB| = |AC| = 2 \) cm. Let \( \theta \) be the angle at the vertex \( A \). If the angle decreases at the rate \( 0.3 \) rad/s, what is the rate of change of the length of the side \( BC \) at the instance when \( \theta = \pi/3 \).
(2) Consider a triangle with vertices $A$, $B$, $C$ such that $|AB| = |AC| = 3$ cm. Let $\theta$ be the angle at the vertex $A$. If the angle increases at the rate $0.3$ rad/s, what is the rate of change of the area of the triangle at the instance when $\theta = \pi/3$.

(3) The sides of a rectangle change with time such that the area of the rectangle does not change. If the rate of change of one side is $1$ m/s, find the rate of change of the other side at the instance when both sides are of equal length.

(4) The sides of a rectangle change with time at the rates $1$ m/s and $-3$ m/s. Find the rate of change of the diagonal of the rectangle at the instance when both sides are of equal length.

(5) At what rate does the area of a disk increase at the instance when the radius is $R = 10$ cm if the radius increases at a constant rate of $2$ cm/s?

(6) At what rate do the area and the length of the diagonal of a rectangle change at the instance when one side is $x = 20$ m and the other is $y = 15$ m if the former is decreasing at a rate of $1$ m/s, while the latter is increasing at a rate of $2$ m/s?

(7) Two ships, $A$ and $B$, leave a harbor at the same time, one heading north and the other heading east. At what rate is the distance between the ships increasing if the speed of ship $A$ is $30$ km/h and the speed of ship $B$ is $40$ km/h?

(8) The surface area of a ball is increasing at a constant rate of $4$ m$^2$/min. At what rate do the radius and volume of the ball change at the instance when the ball has radius $3$ m?

(9) A ladder $24$ ft long leans against a vertical wall. If the lower end is being moved away from the wall at a rate of $3$ ft/sec, how fast is the top descending when the lower end is $8$ ft from the wall? When are the lower and upper ends moving at the same rate?

(10) A man $6$ ft tall walks away from an arc light $15$ ft high at a rate of $3$ miles per hour. How fast is the farther end of his shadow moving? How fast is his shadow lengthening?

(11) The volume of a sphere is increasing at a rate of $16$ cm$^3$/s. How fast is the radius increasing when it is $6$ cm? How fast is the surface area increasing when it is $36$ cm$^2$?

(12) Sand is being poured on the ground from an elevated pipe and forms a pile that always has the shape of a circular cone whose height is equal to the radius of the base. If the sand falls at a rate of $0.5$ m$^3$/min, how fast is the height of the pile increasing when it is $2$ m?
(13) A particle moves along the curve defined by the algebraic equation \( x^2 - 2y^3 = 9 \) so that the coordinate \( x \) increases steadily at a rate of 3 units of length per second. Find the rate of change of the coordinate \( y \) when the particle is at the point \((x, y) = (5, 2)\).

(14) The velocity of a particle moving along a straight line satisfies the condition \( v^2 = c + 2b/s \), where \( a \) and \( b \) are constants and \( s \) is the distance traveled by the particle. Show that the acceleration (the rate of change in velocity with respect to time) is \( a = dv/dt = -b/s^2 \).

(15) Consider two lines \( y + x = 2a \) and \( y - x = 0 \), where \( a \) is a number. Suppose that a particle moves along the first line toward the point of intersection of the lines at a constant speed \( v_1 \), while another particle moves along the second line in the direction away from the point of intersection at a constant speed \( v_2 \). Find the rate of change of the distance between the particles when the first and second particles are at the distances \( s_1 \) and \( s_2 \) from the point of intersection, respectively. In particular, what is the value of this rate if \( s_1 = s_2 \) and \( v_1 = v_2 \)?

(16) The blades of a pair of scissors have width \( 2h \). Find the rate at which the point of intersection of the edges of the blades is moving if the angle between the blades decreases at a constant rate \( \omega \). Assume that the blades are attached by a screw through the midpoint of each blade (i.e., through a point that is at distance \( h \) from the edges of the blade). If \( h = 4 \) mm and \( \omega = -2 \) rad/s, how long should the blades be to see the point of intersection going superluminal?

(17) If \( y^2 = 2x \) and \( x \) is decreasing steadily at a rate of 0.25 units per second, find how fast the slope of the graph is changing at the point \((x, y) = (8, -4)\).

(18) A pool has a spherical bottom of radius \( R \) and the maximal depth \( h < R \) at the pool center. A man walks on the bottom of the pool toward the pool center at a constant speed \( v \). Find the rate at which the man is submerging under the water.

(19) Consider a rectangle with sides \( x \) and \( y \). Suppose that \( x \) and \( y \) change with time. Find their rates of change when \( x = 3 \) cm and \( y = 1 \) cm if, at that moment, the area of the rectangle decreases at a rate of 2 cm\(^2\)/s while the perimeter increases at a rate of 4 cm\(^2\)/s.

(20) Consider a planar region that is a sector of a disk with radius \( R \) and angle \( \varphi \). Suppose that \( R \) and \( \varphi \) change with time so that the area of the region does not change, while its perimeter
increases at a rate of 2 m/s. Find the rate of change of the angle \( \varphi \) and the radius \( R \) at the moment when \( \varphi = 30^\circ \) and \( R = 10 \) m.

21. Linear Approximations and Differentials

21.1. Tangent Line Approximation. The derivative of a function \( f(x) \) at a point \( x = x_0 \) defines the slope of the line tangent to the graph \( y = f(x) \) at the point \((x_0, f(x_0))\) (see Equation (2.5)). The equation of the tangent line is

\[
\frac{y - f(x_0)}{x - x_0} = f'(x_0) \quad \text{or} \quad y = f(x_0) + f'(x_0)(x - x_0).
\]

Definition 3.3. Suppose \( f(x) \) is differentiable at \( x = x_0 \). The linear function

\[
L(x) = f(x_0) + f'(x_0)(x - x_0)
\]

is called the linearization of \( f(x) \) in a neighborhood of \( x_0 \).

Since the values of \( f \) and \( L \) coincide at \( x = x_0 \), one might expect that the difference \( f(x) - L(x) \) is small, provided \( x \) is close enough to \( x_0 \). More precisely, consider the limit

\[
\lim_{x \to x_0} \frac{f(x) - L(x)}{x - x_0} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) = f'(x_0) - f'(x_0) = 0
\]

where the definition of the derivative has been used. This shows that the error of the approximation of \( f(x) \) by \( L(x) \) decreases to zero faster than \( x - x_0 \) as \( x \) approaches \( x_0 \):

\[
f(x) = L(x) + (x - x_0)\varepsilon(x - x_0), \quad \text{where} \quad \varepsilon(x - x_0) \to 0 \quad \text{as} \quad x \to x_0
\]

So the linear function \( L(x) \) may be used to approximate values of \( f(x) \) in a small neighborhood of \( x_0 \). This approximation is called the linear approximation or tangent line approximation. The concept of the tangent line approximation is illustrated in Figure 3.2.

Example 3.12. Use the linear approximation to estimate the value \( \sqrt{3.92} \).

Solution:

(i) Consider \( f(x) = \sqrt{x} \). The closest value of \( x \) to 3.92 at which the square root can be evaluated without a calculator is \( x_0 = 4 \):

\( f(x_0) = 2 \). Note the two important steps here: the choice of \( f(x) \) suitable for the problem and the choice of \( x_0 \) near which the linear approximation is to be used.
2. RULES OF DIFFERENTIATION

### Figure 3.2. Tangent line approximation.

In a neighborhood of \(x_0\) (an interval \( [x_0 - \delta, x_0 + \delta] \)), the tangent line \(y = L(x)\) stays close to the graph \(y = f(x)\). By reducing the width of the interval \(\delta\), one can make the error \(\varepsilon\) of the tangent line approximation as small as desired, i.e., \(|f(x) - L(x)| \leq \varepsilon\) for all \(x \in [x_0 - \delta, x_0 + \delta]\).

(ii) Since \(f'(x) = (\sqrt{x})' = 1/(2\sqrt{x})\) and \(f'(4) = 1/4\), by Equation (3.23) the linearization of \(\sqrt{x}\) near \(x = 4\) is

\[
L(x) = 2 + \frac{1}{4}(x - 4).
\]

(iii) The linear approximation means that the value \(f(3.92) = \sqrt{3.92}\) is approximated by the value \(L(3.92)\):

\[
\sqrt{3.92} \approx L(3.92) = 2 + \frac{1}{4}(3.92 - 4) = 1.98.
\]

A calculator gives \(\sqrt{3.92} \approx 1.9799\). So the approximation error is \(|\sqrt{3.92} - L(3.92)| < 1.02 \cdot 10^{-4}\). It is easy to see that \(L(4.08) = 2.02\) and \(|\sqrt{4.08} - L(4.08)| < 1.02 \cdot 10^{-4}\). In notations given in the caption of Figure 3.2, this observation can be summarized by the following inequality:

\[
|\sqrt{x} - L(x)| < 1.02 \cdot 10^{-4} = \varepsilon \quad \text{if} \quad |x - 4| \leq 0.08 = \delta.
\]

In other words, the values of \(\sqrt{x}\) and its linearization differ by no more than \(1.02 \cdot 10^{-4}\) for all \(3.92 \leq x \leq 4.08\). Naturally, a decrease (increase) in the upper bound for the error would lead to a decrease (increase) in the size of a neighborhood of \(x = 4\) where the linear approximation is accurate.
21.2. Accuracy of the Linear Approximation. The previous example leads to a problem that is extremely important in applications: Given an upper bound for the error \( \varepsilon \) of the linear approximation of a function \( f(x) \) near \( x_0 \), find \( \delta \) such that

\[
|f(x) - L(x)| \leq \varepsilon \quad \text{if} \quad |x - x_0| \leq \delta,
\]
or, alternatively, given \( \delta \), that is, the neighborhood \( x_0 - \delta \leq x \leq x_0 + \delta \), estimate the error \( \varepsilon \) of the linear approximation. The following theorem is useful to answer these questions.

**Theorem 3.15.** Suppose a function \( f(x) \) is twice differentiable in \((a, b)\) such that \( |f''(x)| \leq M \) for all \( x \in (a, b) \) and some number \( M \). Let \( L(x) \) be the linearization of \( f(x) \) at \( x_0 \in (a, b) \). Then

\[
|f(x) - L(x)| \leq \frac{1}{2} M (x - x_0)^2, \quad x \in (a, b).
\]

This theorem is a simpler version of the Taylor theorem, which is proved in advanced calculus courses. The following example illustrates the use of this theorem to assess the accuracy of the linear approximation.

**Example 3.13.** Consider the linearization of \( \sin x \) at \( x = 0 \). Find an interval \( |x| \leq \delta \) in which the error of the linear approximation does not exceed \( \varepsilon = 0.5 \cdot 10^{-3} \).

**Solution:**

(i) Since \( f'(x) = (\sin x)' = \cos x \), \( f'(0) = 1 \), and \( f(0) = 0 \), the linearization is \( L(x) = x \).

(ii) In Theorem 3.15, let \( a = -\delta \) and \( b = \delta \). Next, one has to find \( M \). The simplest way to do this is to take the maximal value of \( |f''(x)| \) in the interval \( |x| \leq \delta \). Note that there should be \( \delta < \pi / 2 \) because \( L(\pi / 2) - \sin(\pi / 2) = \pi / 2 - 1 \) exceeds the given error \( \varepsilon \). So \( \sin x \) is monotonic in \( |x| \leq \delta \), and hence \( |(\sin x)''| = |\sin x| \leq \sin \delta = M \) for all \( |x| \leq \delta \). By Theorem 3.15,

\[
|\sin x - x| \leq \frac{1}{2} M x^2 \leq \frac{1}{2} M \delta^2 = \varepsilon \quad \text{if} \quad |x| \leq \delta.
\]

With \( M = \sin \delta \), the solution of the equation \( \delta^2 \sin \delta = 2\varepsilon = 10^{-3} \) determines \( \delta \). An analytic solution of this equation is impossible. So a value of \( \delta \) has to be found numerically (actually, \( \delta \approx 0.100057 \)).

(iii) Otherwise, one can choose a larger \( M \), for example, \( \sin x \leq 1 \) for any \( x \). So \( M = 1 \) is acceptable, too. This simplifies Equation (3.24): \( \delta^2 = 10^{-3} \) and hence \( \delta \approx 0.0362 \). This value of \( \delta \) appears to be smaller than that in the case \( M = \sin \delta \). It
follows from Equation (3.24) that a larger value of $M$ leads to a smaller $\delta$. So this option should not be “abused.” A good $M$ is not too large and yet is simple enough to solve Equation (3.24). This requires some skills to achieve.

(iv) A good compromise is to use the inequality $\sin \delta \leq \delta$. So the choice $M = \delta$ also fulfills the conditions of Theorem 3.15. Equation (3.24) becomes $\delta^3 = 10^{-3}$ and $\delta = 0.1$, which is to be compared with $\delta = 0.0362$ when $M = 1$ and $\delta \approx 0.100057$ when $M = \sin \delta$.

The converse problem is simpler: Find an upper bound for the error of the linear approximation of $\sin x$ at $x = 0$ in the interval $|x| \leq 0.2$. By monotonicity of $\sin x$ in the interval $(-\pi/2, \pi/2)$, $|\cos x| = |\sin x| \leq \sin(0.2) = M$ for $|x| \leq 0.2$ and, hence, $|\sin x - x| \leq \varepsilon = \frac{1}{2} M \delta^2 = 0.5 \cdot \sin(0.2) \cdot (0.2)^2 \approx 3.9734 \cdot 10^{-3}$.

21.3. Differential. For a real variable $x$, the differential $dx$ is defined as an increment of $x$. It can be given the value of any real number independently of the value of $x$; that is, $dx$ is considered as an independent variable. So, with every real variable, one can associate another real variable, called the differential. If two real variables are related, the following rule postulates the relation between their differentials.

**Definition 3.4.** Let two variables $y$ and $x$ be related as $y = f(x)$, where $f$ is a differentiable function. The differential $dy = df(x)$ is defined by the linear transformation of $dx$:

$$ dy = df(x) = f'(x) \, dx. $$

Note that the variables $x$ and $dx$ on the right-hand side are independent variables. Equation (3.25) states that, if the variables $y$ and $x$ are related, then the differential $dy$ is no longer an independent variable and is determined by $x$ and $dx$; specifically, $dy$ depends linearly on $dx$.

21.4. Geometrical Significance of the Differential. Put $dx = \Delta x$, where $\Delta x$ is a real number. Fix $x = x_0$ and consider an increment of the variable $y = f(x)$ between $x_0 + \Delta x$ and $x_0$:

$$ \Delta y = f(x_0 + \Delta x) - f(x_0) = \Delta f(x_0). $$

The differential $df(x_0) = f'(x_0) \, \Delta x$ does not generally coincide with the increment $\Delta f(x_0)$. For example, put $f(x) = x^2$, $x_0 = 1$, $\Delta x = 0.2$, then $\Delta f(1) = (1 + 0.2)^2 - 1 = 0.44$, whereas $df(1) = f'(1) \, \Delta x = 2 \cdot 0.2 = 0.4$. 
The differential $df(x_0) = f'(x_0)\,dx$ is the increment along the tangent line: $df(x_0) = L(x_0 + \Delta x) - L(x_0)$, $dx = \Delta x$. The differential $df(x_0)$ does not coincide with the increment of the function $\Delta f(x_0) = f(x_0 + \Delta x) - f(x_0) \neq df(x_0)$. Only when $\Delta x$ becomes infinitesimally small, $\Delta x \to 0$, does it coincide up to terms that go to 0 faster than $\Delta x$, i.e., $[\Delta f(x_0) - df(x_0)]/\Delta x \to 0$ as $\Delta x \to 0$.

Since the derivative $f'(x_0)$ determines the slope of the tangent line $L(x) = f(x_0) + f'(x_0)(x - x_0)$ to the graph $y = f(x)$, the differential $df(x_0)$ is the increment of the linearization $y = L(x)$ of the function at $x = x_0$ in the interval $[x_0, x_0 + \Delta x]$, that is, for a particular value $x = x_0$ and an arbitrarily chosen increment $dx = \Delta x$,

$$df(x_0) = L(x_0 + \Delta x) - L(x_0) = f'(x_0)\Delta x.$$ 

Thus, $df(x_0) \neq \Delta f(x_0)$ because the tangent line does not generally coincide with the graph. This observation is summarized in Figure 3.3. In particular, the tangent line approximation can now be stated as

$$f(x_0 + \Delta x) \approx L(x_0 + \Delta x) = f(x_0) + df(x_0), \quad dx = \Delta x.$$

An intuitive understanding of the differential stems from its geometrical interpretation. Let $\Delta x$ tend to 0. The ratio

$$\frac{\Delta y - dy}{\Delta x} = \frac{\Delta f(x) - f'(x)\Delta x}{\Delta x} = \frac{\Delta f(x)}{\Delta x} - f'(x) \to 0$$

converges to zero as $\Delta x \to 0$ because by the existence of $f'(x)$, $\Delta f(x)/\Delta x \to f'(x)$ as $\Delta x \to 0$. This means that the difference $\Delta y - dy$ must go to 0 faster than $\Delta x$. An increment $\Delta x$ is said to be infinitesimally small if $(\Delta x)^n, \, n > 1$, can always be neglected. So one might think of differentials as infinitesimal variations of variables. From this

![Figure 3.3. Geometrical significance of the differential.](image-url)
point of view, the definition (3.25) looks rather natural: Infinitesimal variations of two related variables must be related linearly as their higher powers can always be neglected. The concept of the differential becomes rather practical when one has to establish relations between variations of related quantities in situations when these variations may be viewed as infinitesimal.

21.5. Inverse Function and the Differential. The concept of the differential offers a simple way to find the derivative of an inverse function. Suppose that a function \( f \) has the inverse \( g = f^{-1} \) and \( g \) is differentiable (conditions under which \( g \) exists and is differentiable are stated later in the inverse function theorem of Section 23). If \( y = f(x) \), then the differentials are related as \( dy = f'(x) \, dx \). On the other hand, \( x = f^{-1}(y) = g(y) \) and hence \( dx = g'(y) \, dy \). Since the ratio of the differentials is the derivative, it follows that

\[
\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} \iff g'(f(x)) = \frac{1}{f'(x)} \iff g'(y) = \frac{1}{f'(g(y))}.
\]

For example, \( f(x) = \tan x \) and \( g(y) = \tan^{-1} y \). Then

\[
f'(x) = \frac{1}{\cos^2 x} = 1 + (f(x))^2 \implies g'(y) = \frac{1}{f'(g(y))} = \frac{1}{1 + y^2},
\]

where the relation \( f(g(y)) = y \) has been used.

21.6. Related Errors. Every physical quantity is known only with a certain degree of accuracy. Errors are inherent in the very process of taking measurements. As a point of fact, a value of a physical quantity given without its measurement error does not make much sense; neither should one draw any conclusion from data without a proper analysis of the errors. One of the important practical applications of the differential lies in the error analysis of related quantities.

Suppose there is a relation between two quantities \( y \) and \( x \), \( y = f(x) \). Let \( x \) be measured with an error. This means the following. After \( n \) measurements, one gets \( n \) values \( x_1, x_2, \ldots, x_n \). The average is \( x = (x_1 + x_2 + \cdots + x_n)/n \) is viewed as the actual value. The measured values deviate from the average by amounts \( \Delta x_1 = x_1 - x, \ldots, \Delta x_n = x_n - x \). If \( \Delta x = \max\{|\Delta x_1|, \ldots, |\Delta x_n|\} \) (i.e., \( \Delta x \) is the maximal of the absolute errors), then all measured values lie in the interval \([x - \Delta x, x + \Delta x]\). The quantity \( \Delta x \) is the maximum uncertainty in the value of \( x \) (or an error bound). One writes for the measured value \( x \pm \Delta x \) to indicate the average measured value and its maximum uncertainty. The number
Δx is usually known and determined by the very process of taking measurements.

A standard question in error analysis: What is the accuracy of the value y = f(x)? Apparently, x and Δx are independent variables as the error bound Δx depends on the way in which the variable is measured (there are more and less accurate methods which would lead to smaller and higher values of Δx independently of the value of x). Naturally, one might assume that the errors are small; that is, they are infinitesimal variations of measured quantities. Then the errors of the related quantities must be related as their differentials. This is a standard assumption of the error analysis. In other words, if y = f(x) where x is the measured mean value, then the error in the value of y in each measurement is assumed to be

\[ \Delta y_i = y_i - y = f(x_i) - f(x) = f(x + \Delta x_i) - f(x) \]

which is nothing but dy = df(x) where dy = Δy_i if dx = Δx_i. The absolute value of the differential |dy| represents an absolute error of y = f(x). The quantity |dy/y| · 100% is called a relative error. The absolute and relative error bounds are, respectively

\[ \Delta y = |f'(x)|\Delta x, \quad \frac{\Delta y}{|y|} \cdot 100% = \frac{|f'(x)|}{|f(x)|} \Delta x \cdot 100\% \]

and one writes for the measured value y ± Δy to indicate the maximum uncertainty in the value of y.

Example 3.14. What are the absolute and relative error bounds of the volume of a cube if its side is 10 ± 0.1 cm?

Remark. When measuring the length by a ruler with a grid, the measurement error should not exceed the ruler grid spacing (e.g., a ruler with a millimeter grid).

Solution: The volume V and side x are related as V = x³. So dV = 3x²dx. Setting dx = 0.1 cm and x = 10 cm, dV = 30 cm³ and V = 1000 ± 30 cm³. The relative error bound is dV/V = 0.03 or 3% (note that dx/x = 0.01, i.e., only 1%).

The error analysis for several related quantities is studied in multivariable calculus courses. It is based on the concept of the differential of functions of several variables.
21.7. Exercises.

(1) Find the linearizations of each of the following functions at the specified point:
   (i) \( \cos x, \ x = \pi/4 \)
   (ii) \( \tan x, \ x = 0 \)
   (iii) \( e^{-x^2}, \ x = 0 \)
   (iv) \( \ln x, \ x = e \)
   (v) \( \sqrt{1+x}, \ x = 3 \)

(2) Estimate the error of the tangent line approximation of each of the following functions over an interval \( |x - x_0| \leq \delta \) for the specified point \( x_0 \) and the width \( \delta \):
   (i) \( \sqrt{1+x}, \ x_0 = 3, \ \delta = 0.1 \)
   (ii) \( \ln x, \ x = 1, \ \delta = 0.2 \)
   (iii) \( \tan x, \ x = 0, \ \delta = \pi/4 \)

(3) Find the differentials of each of the following functions:
   (i) \( x(1-x^2)^3 \)
   (ii) \( (y-2)/(y+1) \)
   (iii) \( \sqrt{1+x^2}/x \)
   (iv) \( \sin^2 t + \cos(t^2) \)
   (v) \( \ln(x+2) + xe^x \)
   (vi) \( a^{-1} \tan^{-1}(x/a), \ a \neq 0 \)
   (vii) \( \ln |x + \sqrt{x^2 + a}| \)
   (viii) \( (2a)^{-1} \ln |(x-a)/(x+a)| \)

(4) Use differentials (the tangent line approximation) to estimate the following numbers and assess the accuracy of the estimates:
   (i) \( \sqrt{24.6} \)
   (ii) \( e^{0.08} \)
   (iii) \( \sqrt{1.02} \)
   (iv) \( \sin 29^\circ \)
   (v) \( \tan^{-1} 1.05 \)
   (vi) \( \log_{10} 11 \)

(5) Prove the approximation formula
   \[
   \sqrt[n]{a^n + x} \approx a + \frac{x}{n a^{n-1}}, \quad a > 0
   \]
   and use it to calculate approximately the numbers
   (i) \( \sqrt[3]{9} \)
   (ii) \( \sqrt[4]{80} \)
   (iii) \( \sqrt[5]{100} \)
   (iv) \( \sqrt[10]{1000} \)
(6) Calculate $\Delta f(1)$ and $df(1)$ for the function $f(x) = x^3 - 2x + 1$ and compare them in the following three cases: $\Delta x = 1$, $\Delta x = 0.1$, and $\Delta x = 0.01$.

(7) Let $u$, $v$, and $w$ be differentiable functions. Find $dy$ if
(i) $y = uvw$
(ii) $y = u/v^2$
(iii) $y = (u^2 + v^2)^{-1/2}$
(iv) $y = \ln(\sqrt{u^2 + v^2})$
(v) $y = \tan^{-1}(u/v)$

(8) Find $dy$ in terms of $x$, $y$, and $dx$ if
(i) $\sqrt{x} + \sqrt{y} = 4$
(ii) $y^3 + x^3 = 2xy$
(iii) $\cos(x + 3y) = \sin(xy)$

(9) Find an approximate formula for the area of a circular ring of radius $r$ and width $dr$. What is the exact formula?

(10) Find an approximate formula for the volume of a spherical shell of radius $r$ and thickness $dr$. Assess the accuracy of the approximation by stating the condition on $r$ and $dr$ so that the relative error does not exceed $\varepsilon = 0.01$ (i.e., 1%).

(11) What is an admissible relative error in measurements of the radius of a ball in order for the relative error of the volume to be less than 1%?

(12) A sector of a disk of radius $R = 100$ cm has an angle $\theta = 60^\circ$. How much is the area of the sector changed if
(i) the radius $R$ is increased by 1 cm?
(ii) the angle $\theta$ is decreased by 30$'$?
Give the exact and approximate solutions. Compare them.

(13) The period of a pendulum is determined by the equation

$$T = 2\pi \sqrt{\frac{l}{g}},$$

where $l$ is the length of the pendulum (in cm) and $g = 981$ cm/s$^2$ is the free-fall acceleration. How much should the length $l = 20$ cm be changed in order to increase the period by 0.05 s?

(14) To determine the free-fall acceleration, the period of a pendulum is measured so that by the above equation $g = 4\pi^2l^2/T^2$. How do the measured values of $g$ vary if
(i) $T$ is measured with a relative error bound $\varepsilon$?
(ii) $L$ is measured with a relative error bound $\varepsilon$?
(15) Find the absolute error of \( \log_{10} x \) \((x > 0)\) if the relative error of \( x \) is \( \varepsilon \).

(16) Use differentials to find the derivatives of the inverse functions \( \sin^{-1} x \) and \( \cos^{-1} x \).

(17) Prove that the linearization of a differentiable function \( f(x) \) in a neighborhood of \( x_0 \) is unique in the sense that if \( L(x) = b + m(x - x_0) \) and
\[
\lim_{x \to x_0} \frac{f(x) - L(x)}{x - x_0} = 0,
\]
then \( b = f(x_0) \) and \( m = f'(x_0) \). In other words, the linearization is the only linear approximation whose error decreases to zero faster than \( x - x_0 \) as \( x \) approaches \( x_0 \).

(18) Find the tangent line and the normal line (the line perpendicular to the tangent line) to the curve \( y = (x + 1)\sqrt{3 - x} \) at the points \((-1, 0)\), \((2, 3)\), and \((3, 0)\).

(19) Find the point(s) of the parabola \( y = 2 + x - x^2 \) at which the tangent line is (i) parallel to the \( x \) axis and (ii) parallel to the line \( y = x \).

(20) Prove the relation
\[
\sqrt{a^2 + x} = a + x + R, \quad 0 < R < \frac{x^2}{8a^3},
\]
where \( a > 0 \) and \( x > 0 \).
CHAPTER 4

Applications of Differentiation

22. Minimum and Maximum Values

Some of the most important applications of calculus are optimization problems. An example of an ancient optimization problem: A man can throw a stone at a speed of $v_0$. At what angle should the stone be thrown in order to get the maximal range? An example of a modern optimization problem: How can one optimize the information flow in the World Wide Web to avoid crashes of servers? Many of these problems can be reduced to finding the maximal and minimal values of a given function.

**Definition 4.1 (Absolute Maximum and Minimum).** A function $f$ has an absolute maximum at $c$ if $f(x) \leq f(c)$ for all $x$ in the domain $D$ of $f$. Similarly, the value $f(c)$ is called the maximum value of $f$. A function $f$ has an absolute minimum at $c$ if $f(x) \geq f(c)$ for all $x$ in the domain $D$ of $f$. The value $f(c)$ is called the minimum value of $f$. The maximum and minimum values of $f$ are called the extreme values of $f$.

For example, the function $f(x) = \cos x$ attains its maximum value 1 at $x = 2\pi n$, where $n = 0, \pm 1, \pm 2, \ldots$, and its minimum value $-1$ at $x = \pi + 2\pi n$. A function does not always have a maximum or minimum value. For instance, the function $f(x) = 1/x$ defined for all real $x \neq 0$ has neither maximum nor minimum value because, for any real $M$, one can always find $x$ such that $f(x) > M$ ($0 < x < 1/|M|$). No real number can be the maximum value of $f(x)$. Similarly, for any real $M$, $f(x) < M$ if $-1/|M| < x < 0$; that is, no minimum value exists. The function $f(x) = x^2$ has no maximum value on the real axis, but it does have an absolute minimum at $x = 0$ because $x^2 \geq 0$ for all $x$ and $f(0) = 0$, that is, $f(x) \geq f(0)$.

22.1. Relative Maxima and Minima.

**Definition 4.2 (Local Maximum and Minimum).** A function $f$ has a local (or relative) maximum at $c$ if $f(x) \leq f(c)$ for all $x$ in some open interval containing $c$. Similarly, a function $f$ has a local (or
relative) minimum at } \( c \) if } \( f(x) \geq f(c) \) for all } \( x \) in some open interval containing } \( c \).

**Example 4.1.** Does the function } \( f(x) = x^3 - x = x(x^2 - 1) \) have an absolute maximum (minimum) value and relative maxima (minima) on the real axis?

![Graph of the function](image)

**Figure 4.1.** Graph of the function } \( f(x) = x^3 - x = x(x^2 - 1) \). It does not have an absolute maximum or minimum value. However, it does have a relative maximum at } \( x = -1/\sqrt{3} \) and a relative minimum at } \( x = 1/\sqrt{3} \).

**Solution:**

1. The function has neither an absolute maximum nor an absolute minimum because it grows unboundedly with increasing } \( x \) and it decreases unboundedly as } \( x \) attains larger negative values.
2. The function vanishes at three points } \( x = 0, \pm 1 \). It can have relative minima and maxima between its zeros because the values of } \( f \) are bounded from above and below: } \(|f(x)| \leq |x|^3 + |x| \leq 2\) for } \(|x| \leq 1\), that is, } \(-2 \leq f(x) \leq 2\) if } \(-1 \leq x \leq 1\).
3. Consider the open interval } \( x \in (0,1) \). The function is strictly negative in it and bounded from below: } \( M < f(x) < 0 \) for all } \( x \in (0,1) \) (e.g., } \( M = -3\)). By increasing } \( M \), one can eventually reach the situation when there is } \( 0 < c < 1 \) such that } \( M = f(c) \leq f(x) \) for all } \( 0 < x < 1 \). This happens when the horizontal line } \( y = M \) touches the graph } \( y = f(x) \). Thus, } \( f \) must have a relative minimum in } \( (0,1) \).

**Remark.** The actual value } \( c = 1/\sqrt{3} \). How is it obtained? There is a technique to find } \( c \), which will be studied shortly.

4. Similarly, } \( f \) is strictly positive in } \( (-1,0) \) and bounded from above } \( 0 < f(x) < M \) for some } \( M \). By lowering the horizontal line } \( y = M \) (or decreasing } \( M \)) to the point when it touches the graph } \( y = f(x) \), one
can find a point \( c \in (-1, 0) \) such that \( f(x) \leq f(c) \) for all \( x \in (-1, 0) \); that is, \( f \) has a relative maximum in \((-1, 0)\).

Remark. The actual value is \( c = -1/\sqrt{3} \) (see below).

One of the lessons that can be learned from this example is that one can think of a relative minimum (maximum) as an absolute minimum (maximum) when \( f \) is restricted to a sufficiently small subset in its domain. This observation is accurately stated by the following theorem.

**Theorem 4.1 (The Extreme Value Theorem).** If \( f \) is a continuous function on a closed interval \([a, b]\), then \( f \) attains its absolute maximum and minimum values in \([a, b]\); that is, there exist \( c_1 \) and \( c_2 \) in \([a, b]\) such that \( f(c_1) \leq f(x) \leq f(c_2) \) for all \( x \) in \([a, b]\).

![Figure 4.2. Extreme value theorem. An example of a continuous function with several local minima and maxima. The minimal value coincides with one of the local minima, while the maximal value is reached at the endpoint of the interval: \( f(c_1) \leq f(x) \leq f(b) \) for all \( x \in [a, b] \). The hypothesis of the closedness of the interval is crucial. If the point \( b \) is excluded, then \( f \) has no maximal value on \([a, b]\).](image)

The continuity hypothesis is essential. In fact, the continuity of \( f(x) = x^3 - x \) was implicitly used in Example 4.1 to establish the existence of its relative maximum and minimum! The following example illustrates the point. Consider the function \( f(x) = 2x \) if \( x \in [0, 1) \) and \( f(x) = 1 \) if \( x \in [1, 2] \). So the function is defined on the closed
interval $[0, 2]$ and bounded from above $f(x) < M$ (e.g., $M > 2$). An attempt to establish the existence of a maximum value of $f$ by lowering $M$ fails! Indeed, the lowest upper bound is $M = 2$, but there is no $c$ such that $f(c) = 2$. The values of $f$ approach $2$ as $x$ approaches $1$ from the left, but $f(1) = 1$! For any positive $\epsilon > 0$, $f(1 - \epsilon) < f(x)$ for $x \in (1 - \epsilon, 1)$ no matter how small $\epsilon$ is. Thus, $f$ does not have an absolute maximum value because of its discontinuity at $x = 1$. The absolute minimum exists: $f(0) \leq f(x)$. Note that the function $f(x) = 2x$ when $x \in [0, 1]$ and $f(x) = 1$ when $x \in (1, 2]$ has an absolute maximum and minimum, $f(0) \leq f(x) \leq f(1)$, despite its discontinuity at $x = 1$. So the continuity hypothesis is a sufficient condition, but not necessary.

The hypothesis of the closedness of the interval is also a sufficient condition, but not necessary. The continuous function $f(x) = x$ does not attain its absolute maximum or minimum value on any open interval $(a, b)$. But it does so if the interval becomes closed: $f(a) \leq f(x) \leq f(b)$ for any $x \in [a, b]$. On the other hand, the continuous function $f(x) = x^3 - x$ in the open interval $(-1, 1)$ attains its absolute maximum and minimum value as one can see in Figure 4.1.

22.2. Derivatives at Local Maxima and Minima. The second observation resulting from Example 4.1 is that at the point where a continuous function attains its local minimum or maximum value there is a horizontal line that touches the graph of this function. So, if, in addition, the function is differentiable, then this horizontal line is a tangent line with the vanishing slope; that is, the derivative of the function vanishes at points where the function attains its local maximum or minimum value.

**THEOREM 4.2** (Fermat’s Theorem). If $f$ has a local maximum or minimum at $c$, and if $f'(c)$ exists, then $f'(c) = 0$.

**PROOF.** By the existence of $f'(c)$

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

Therefore, the right and left limits must coincide with $f'(c)$ (see Section 7.3):

$$(4.1) \quad \lim_{h \to 0^-} \frac{f(c + h) - f(c)}{h} = f'(c) = \lim_{h \to 0^+} \frac{f(c + h) - f(c)}{h}.$$ 

Let $f$ have a local maximum (the case of a local minimum can be treated similarly). Then $f(c) \geq f(x)$ or $f(x) - f(c) \leq 0$ in some open interval $a < x < b$. In particular, $[f(c + h) - f(c)]/h \leq 0$ for any
positive $h > 0$ such that $c < c + h < b$. By Theorem 2.2,

\[(4.2) \quad \frac{f(c + h) - f(c)}{h} \leq 0 \implies \lim_{h \to 0^+} \frac{f(c + h) - f(c)}{h} \leq 0.\]

Similarly, for any negative $h < 0$ such that $a < c + h < c$, one has $\lfloor f(c + h) - f(c) \rfloor \leq 0$ and $\lfloor f(c + h) - f(c) \rfloor/h \geq 0$. Hence,

\[(4.3) \quad 0 \leq \frac{f(c + h) - f(c)}{h} \implies 0 \leq \lim_{h \to 0^-} \frac{f(c + h) - f(c)}{h}.\]

By inequalities (4.2) and (4.3), it follows from (4.1) that

\[0 \leq f'(c) \leq 0,\]

which is only possible if $f'(c) = 0$. □

This theorem provides a powerful tool to determine the actual positions of local maxima and minima. Let us go back to Example 4.1 ($f(x) = x^3 - x$). The slope $f'(x) = 3x^2 - 1$ vanishes at two points $x = \pm 1/\sqrt{3}$. According to the analysis carried out in Example 4.1, $f$ has a local maximum at $x = -1/\sqrt{3} \in (-1, 0)$ and a local minimum at $x = 1/\sqrt{3} \in (0, 1)$.

**Definition 4.3.** A number $c$ in the domain of a function $f$ is said to be a critical point of $f$ if either $f'(c) = 0$ or $f'(c)$ does not exist.

Does the equation $f'(x) = 0$ determine all local maxima and minima of $f$?

(I) A function may have a local minimum or maximum at a point where the derivative does not exist. A simple example is the function $f(x) = |x|$. It has an absolute minimum at $x = 0$, but $f'(x)$ does not exist at $x = 0$. So this minimum cannot be found from $f'(x) = 0$.

(II) If $f$ is differentiable everywhere, then, by solving $f'(x) = 0$, all local minima and maxima can be found. However, not all the solutions generally correspond to either a local maximum or a local minimum. The function $f(x) = x^3$ has no minimum or maximum, but its derivative $f'(x) = 3x^2$ vanishes at $x = 0$. In other words, the converse of Fermat’s theorem is false.

(III) If all critical points of a function are found, then their type (local maximum, local minimum, or none of the above) can be analyzed by comparing values $f(c \pm h)$ with $f(c)$, where $c$ is a critical point (cf. Definition 4.2). If $f''(c)$ exists, then the second derivative test can be used, which is discussed later.
(IV) A function defined on a closed interval \([a, b]\) can have its absolute maximum or minimum at the endpoints. When finding the absolute maximum and minimum values, the values of \(f\) at the critical points must be compared with \(f(a)\) and \(f(b)\). The largest (smallest) of them is the absolute maximum (minimum) value.

**Example 4.2.** If a stone is thrown at a speed \(v_0\) m/s and an angle \(\theta\) with the horizontal line, then its trajectory is a parabola:

\[
y = x \tan \theta - \frac{x^2 g}{2v_0^2 \cos^2 \theta},
\]

where \(y\) is the stone height (vertical position), \(x\) is the horizontal position (all the positions are in meters), and \(g = 9.8\) m/s\(^2\) is a constant universal for all objects near the surface of the Earth (the free-fall acceleration). This is a consequence of the Newton’s second law. At what angle should one throw a stone to reach the maximal range at a given speed \(v_0\)?

**Solution:** 1. The range as a function of the angle \(\theta\) has to be found first. The stone lands when its height \(y\) vanishes. The equation \(y = 0\) has two solutions \(x = 0\) (naturally, this is where the stone was thrown) and \(x = L(\theta)\), where

\[
L(\theta) = \frac{2v_0^2}{g} \tan \theta \cos^2 \theta = \frac{2v_0^2}{g} \sin \theta \cos \theta = \frac{v_0^2}{g} \sin(2\theta).
\]

2. The range \(L(\theta)\) is a differentiable function of \(\theta\) so the values of \(\theta\) at which \(L\) attains its extreme values may be found from the equation

\[
L'(\theta) = 0 \quad \implies \quad \frac{v_0^2}{g} 2 \cos(2\theta) = 0 \quad \implies \quad \cos(2\theta) = 0.
\]

This equation has countably many solutions \(2\theta = \pi/2 + \pi n\), where \(n\) is any integer. But in the interval of the physical values of \(\theta \in [0, \pi/2]\), it has only one solution \(\theta = \pi/4\). Since \(\sin(2\pi/4) = 1\) (the absolute maximum of the sine), \(L\) attains its maximum value at \(\theta = \pi/4\). So the range is maximal, \(L_{\text{max}} = \frac{v_0^2}{g}\), when a stone is thrown at 45°.

**Remark.** The conclusion in the preceding example is independent of the stone’s mass and its initial speed \(v_0\). In reality, for larger values of \(v_0\), like a projectile shot by a gun, trajectory would deviate from the parabola (due to friction with the air that increases with increasing the speed). So the optimal angle would deviate from \(\pi/4\). The deviation would also depend on the mass and the initial speed. The range
optimization problem becomes more involved and would require the theory of differential equations. It should also be noted that the angle at which the maximal range is attained depends on the initial height at which the stone is thrown. So the angle would be different from 45° when, for example, the stone is thrown from a cliff.

22.3. Exercises.

(1) Examine the following functions for maxima and minima. Draw the graph in each case.
   (i) \( y = 2 + x - x^2 \)
   (ii) \( y = (x - 1)^3 \)
   (iii) \( y = (x + 1)^4 \)
   (iv) \( y = x^2 - 5x + 3 \)
   (v) \( y = 2x^3 - 3x^2 + 6x - 3 \)
   (vi) \( y = x^2 + 16/x \)
   (vii) \( y = x^2 - 1/x^2 \)
   (viii) \( y = 4x/(x^2 + 1) \)
   (ix) \( y = \sin x + \cos x \)
   (x) \( y = xe^x \)
   (xi) \( y = x^n(1 - x)^m \), where \( n \) and \( m \) are positive integers
   (xii) \( y = x^{1/3}(1 - x)^{2/3} \)

(2) Find all critical points of the following functions and determine whether there is a local maximum, a local minimum, or none of the above at each critical point.
   (i) \( y = |x - 3| \)
   (ii) \( y = |x^2 - 4| + 2x \)
   (iii) \( y = |x^3 - 1| \)
   (iv) \( y = |\sin(2x)| \)
   (v) \( y = |x^3|e^x \)
   (vi) \( y = (x - 1)^{1/3} \)
   (vii) \( y = x(x + 1)^{2/3} \)
   (viii) \( y = (1 - x^2)^{3/2} \)

(3) Find the extreme values of the following functions on the specified interval or show that such values do not exist.
   (i) \( y = x^4 - 4x^2, -3 < x < 3 \)
   (ii) \( y = x^4 - 4x^2, -3 \leq x \leq 3 \)
   (iii) \( y = e^x + e^{-x}, -\infty < x < \infty \)
   (iv) \( y = x^3, -1 < x \leq 2 \)
   (v) \( y = \frac{1}{x} + \frac{1}{1-x}, 0 < x < 1 \)

(4) On the circle given by the equation \( x^2 + y^2 = 25 \), find the point nearest to the point \((6,8)\).
5. A line is drawn through a point \((a, b)\) such that the part intercepted between the axes has a minimum length. Prove that the minimum length is \((a^{2/3} + b^{2/3})^{3/2}\).

6. Find the maximum area of an isosceles triangle with fixed perimeter \(p\).

7. Let the sum of two numbers be \(s\). Find the numbers in each of the following cases:
   (i) The sum of their squares is a minimum.
   (ii) The sum of their cubes is a minimum.
   (iii) Their product is a maximum.
   (iv) The difference between one and the reciprocal of the other is a maximum.

8. Can one claim that if the function \(f(x)\) has a maximum \(x = x_0\), then in a sufficiently small neighborhood to the left of \(x_0\) the function \(f(x)\) increases and in a sufficiently small neighborhood to the right of \(x_0\) it decreases? Consider the example:

   \[
   f(x) = 2 - x^2 \left(2 + \sin \left(\frac{1}{x}\right)\right) \quad \text{if} \ x \neq 0 \quad \text{and} \quad f(0) = 2.
   \]

9. Does the function

   \[
   f(x) = |x| \left(2 + \cos \left(\frac{1}{x}\right)\right) \quad \text{if} \ x \neq 0 \quad \text{and} \quad f(0) = 0
   \]

   have a local extreme value at \(x = 0\)? Graph the function.

23. The Mean Value Theorem

**Theorem 4.3 (Rolle’s Theorem).** Let \(f\) be a function that satisfies the following three hypotheses:

(I) \(f\) is continuous on the closed interval \([a, b]\).
(II) \(f\) is differentiable on the open interval \((a, b)\).
(III) \(f(a) = f(b)\).

Then there is a number \(c\) in \((a, b)\) such that \(f'(c) = 0\).

This theorem provides a useful method to prove the existence of a local maximum or minimum of a function \(f\) when analytic solutions of the equation \(f'(x) = 0\) are hard to find. In fact, it has already been used in Example 4.1: The function \(f(x) = x^3 - x\) on the intervals \([-1, 0], [0, 1], [-1, 1]\) satisfies the hypotheses of Rolle’s theorem because \(f(\pm 1) = f(0) = 0\). The proof follows closely the arguments of Example 4.1.
Figure 4.3. Rolle’s theorem. The continuity of \( f \) guarantees the boundedness of \( f \). So the graph of \( f \) lies between two horizontal lines. By lowering an upper bound or increasing a lower bound until one of the horizontal lines (or both) touches the graph and becomes its tangent line, differentiability of \( f \) ensures the existence of the tangent line at every point in \((a, b)\). The slope of the horizontal tangent line is 0 and so is the derivative at that point.

**Proof of Theorem 4.3.**

1. If \( f(x) = f(a) = k \) is a constant function, then \( f'(x) = 0 \) everywhere.

2. Let \( f(x) > f(a) \) for some \( x \in (a, b) \) (cf. Example 4.1 for \( x \in [-1, 0] \)). Since \( f \) is continuous, the extreme value theorem applies, and therefore \( f \) has a maximum in \([a, b]\). Since \( f(a) = f(b) \), the maximal value must be attained at \( c \in (a, b) \). By Fermat’s theorem, \( f'(c) = 0 \) because \( f \) is differentiable in \((a, b)\).

3. If \( f(x) < f(a) \) for some \( x \in (a, b) \) (cf. Example 4.1 for \( x \in [-1, 1] \) or \( x \in [0, 1] \)), then, by the extreme value theorem, \( f \) has a minimum at \( c \in (a, b) \), and, by Fermat’s theorem, \( f'(c) = 0 \).

Rolle’s theorem is also useful to analyze the root pattern of a function.

**Example 4.3. How many real roots does the equation \( x^5 + x^3 + x - 1 = 0 \) have?**

**Solution:** 1. Let \( f(x) = x^5 + x^3 - 1 \). Evidently, \( f(-1) = -4 < 0 \) and \( f(1) = 2 > 0 \). By continuity, \( f \) has to take all intermediate values
between $-\frac{1}{2}$ and $2$ (the intermediate value theorem). So $f$ has at least one root in $(-1, 1)$.

2. Suppose it has two roots $a$ and $b$, that is, $f(a) = f(b) = 0$. Then, by Rolle’s theorem, $f'(x)$ has to vanish somewhere in $(a, b)$. But this is not possible because $f'(x) = 5x^4 + 3x^2 + 1 > 0$ for any $x$. Thus, $f$ has the only real root. □

**Theorem 4.4 (The Mean Value Theorem).** Let $f$ be a function that satisfies the following hypotheses:

(I) $f$ is continuous on the closed interval $[a, b]$.

(II) $f$ is differentiable on the open interval $(a, b)$.

Then there is a number $c \in (a, b)$ such that

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

The geometrical interpretation of the theorem is simple (see Figure 4.4). Consider the line through the points $(a, f(a))$ and $(b, f(b))$. Its slope is $(f(b) - f(a))/(b - a)$. The theorem asserts the existence of a point where the graph $y = f(x)$ has a tangent line with the same slope (cf. Equation (4.5)) (as $f'(c)$ is the slope of the tangent line at $x = c$). Let us turn to a formal proof.

**Proof of Theorem 4.4.**

1. Consider the line through the points $(a, f(a))$ and $(b, f(b))$. Its equation is

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

$$L(a) = f(a), \quad L(b) = f(b).$$

Next, consider the function

$$h(x) = f(x) - L(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a} (x - a).$$

Its values determine the deviation of the graph $y = f(x)$ from the secant line $y = L(x)$ on the closed interval $[a, b]$.

2. The function $h(x)$ satisfies the three hypotheses of Rolle’s theorem. First, it is continuous on $[a, b]$ as the sum of two continuous functions $f(x)$ and $-L(x)$ (a linear function is continuous). Second, it is differentiable on $(a, b)$ as the sum of two differentiable functions:

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}, \quad x \in (a, b).$$
23. THE MEAN VALUE THEOREM

Figure 4.4. Mean value theorem. The secant line of the graph of \( f \) through the points \((a, f(a))\) and \((b, f(b))\) has the slope \( \tan \alpha = \frac{f(b) - f(a)}{b - a} \), where \( \alpha \) is the angle between the secant line and the horizontal line. If \( f \) does not coincide with the secant line, then near \( x = a \) the slope of the tangent line does not coincide with \( \tan \alpha \). Here the case when this slope is greater than \( \tan \alpha \) is shown. Then the graph of \( f \) lies above the secant line near \( x = a \). But the graph has to return to the secant line again. Near the point where the graph and the secant lines meet again, the tangent line has to have a smaller slope than \( \tan \alpha \). So at some point \( c \) the tangent line has to be parallel to the secant line, meaning that \( f'(c) = \tan \alpha \).

Finally, by (4.6) and (4.7), \( h(a) = f(a) - L(a) = 0 \) and \( h(b) = f(b) - L(b) = 0 \), that is, \( h(a) = h(b) \).

3. By Rolle’s theorem, there is a number \( c \in (a, b) \) such that

\[
h'(c) = 0 \implies f'(c) = \frac{f(b) - f(a)}{b - a},
\]

where Equation (4.8) has been used.

Example 4.4. A speeding car was pulled over on an interstate road and a state trooper gave a warning to the driver. Forty five minutes later and passed 65 miles on the road, the car stopped at a rest area.
Another state trooper approached the driver and issued a speeding ticket, claiming that the driver ignored the warning and exceeded 86 miles per hour. Was the trooper’s claim correct?

Solution: Let \( s(t) \) be the distance traveled by the car after it was pulled over the first time. The rate of change \( s'(t) = v(t) \) is the speed of the car at any moment of time. The function \( s(t) \) is defined between \( t = 0 \) and \( t = 45 \text{ min} = 0.75 \text{ hr} \) so that \( s(0) = 0 \) and \( s(0.75) = 65 \text{ mi.} \) It is differentiable as \( s'(t) \) is the car speed! By the mean value theorem, there is a time moment \( t = c \in (0, 0.75) \) when

\[
s'(c) = v(c) = \frac{s(0.75) - s(0)}{0.75 - 0} = \frac{65}{0.75} \approx 86.7 \text{ mi/hr}.
\]

The speeding ticket is justified.

For any two moments of time \( a \) and \( b \), the ratio \( (s(b) - s(a))/(b - a) \) is the average speed on the time interval \( [a, b] \). The mean value theorem simply states that a moving object always attains its average speed at least at one moment of time between \( a \) and \( b \). So, if at time moment \( b \) the object appears to be traveling slower than its average speed, prior to that it must have been traveling faster than its average speed.

Example 4.5. Suppose the derivative \( f' \) exists and is bounded on \((a, b)\), that is, \( m \leq f'(x) \leq M \). If \( f(a) \) is given, how small and how large can \( f(b) \) possibly be?

Solution: By the mean value theorem, there is a \( c \in (a, b) \) such that

\[
f(b) = f(a) + f'(c)(b - a) \leq f(a) + M(b - a).
\]

This equation is easy to understand with the help of a mechanical analogy: How far can a car travel in time \( b - a \) if its speed is not lower than \( m \), but cannot exceed \( M \)?

23.1. Properties of the First Derivative. The derivative of a constant function vanishes. How about the converse? The following theorem answers this question.

Theorem 4.5. If \( f'(x) = 0 \) for all \( x \) in an interval \( (a, b) \), then \( f \) is constant on \( (a, b) \).

Proof. Take any two numbers \( x_1 \) and \( x_2 \) between \( a \) and \( b \). By the mean value theorem, there is a number \( c \) between \( x_1 \) and \( x_2 \) such that \( f(x_1) - f(x_2) = f'(c)(x_1 - x_2) \). By hypothesis, \( f'(c) = 0 \) for any \( c \).
Thus, \( f(x_1) - f(x_2) = 0 \) or \( f(x_1) = f(x_2) \) for any \( x_1 \) and \( x_2 \) in \((a, b)\); that is, \( f \) is constant. \( \square \)

The hypothesis that \( f'(x) = 0 \) in a single interval is crucial. For example, the sign function \( f(x) = 1 \) if \( x > 0 \), and \( f(x) = -1 \) if \( x < 0 \), has zero derivative at any point of its domain, but it is not constant. The key point to note is that the domain is not a single interval, but a union of two disjoint intervals \((-\infty, 0)\) and \((0, \infty)\). So the mean value theorem is not applicable to any interval containing \( x = 0 \). This example is easily extended to the case when the domain is any collection of disjoint intervals and \( f \) takes different constant values on different intervals.

**Corollary 4.1.** If \( f'(x) = g'(x) \) for all \( x \) in an interval \((a, b)\), then \( f - g \) is constant, that is, \( f(x) = g(x) + k \), where \( k \) is a constant.

**Proof.** Let \( h(x) = f(x) - g(x) \). Since \( h'(x) = f'(x) - g'(x) = 0 \) in \((a, b)\), \( h \) is constant, and the conclusion follows. \( \square \)

The sign of the first derivative defines intervals of growth and decrease of a function.

**Theorem 4.6 (Increasing-Decreasing Test).**

(I) If \( f' > 0 \) on an interval, then \( f \) is increasing on that interval.

(II) If \( f' < 0 \) on an interval, then \( f \) is decreasing on that interval.

**Proof.** Take any two numbers \( x_1 \) and \( x_2 \) in the interval so that \( x_1 < x_2 \). A function is increasing if \( f(x_1) < f(x_2) \) and decreasing if \( f(x_1) > f(x_2) \). Since \( f \) is differentiable, the mean value theorem states that there is a number \( c \) between \( x_1 \) and \( x_2 \) such that

\[
(4.9) \quad f(x_2) - f(x_1) = f'(c)(x_2 - x_1).
\]

If \( f' > 0 \), then it follows from (4.9) that \( f(x_2) - f(x_1) > 0 \) because, by assumption, \( x_2 > x_1 \); that is, the function is increasing. Similarly, for \( f' < 0 \), \( f(x_2) - f(x_1) < 0 \), and the function is decreasing. \( \square \)

The function \( f \) is said to be **monotonic** in an interval if it is increasing or decreasing in this interval. By the increasing-decreasing test, the function is monotonic in an interval if its derivative \( f' \) does not vanish in the interval.

The increasing-decreasing test is further illustrated on the interactive website at http://www.math.ufl.edu/~mathguy/ufcalcbook/inc_dec.html.
23.2. The Inverse Function Theorem. A Baby Version. Given a function $f$, its inverse function exists if $f$ is one-to-one as explained in Section 5. A simple rule to calculate the derivative of the inverse function was presented in Sections 17 and 21. However, the very question of whether the inverse function is actually differentiable has not been addressed. It appears that if $f$ is differentiable, then the questions about the existence of the inverse function $f^{-1}$ and its differentiability can be answered by looking at the sign of the derivative $f'$.

**Theorem 4.7.** (A Baby Version of the Inverse Function Theorem). Let $f$ be a function on $-\infty \leq a < b \leq \infty$. Suppose that $f'(x) > 0$ (or $f'(x) < 0$) for all $x \in (a, b)$. Then $f$ has the inverse $g = f^{-1}$ on $(c, d)$ for some $-\infty \leq c < d \leq \infty$ and


g'(f(x)) = \frac{1}{f'(x)}, \quad a < x < b.

**Proof.**

1. Let $f'(x) > 0$. The other case is similar. By the increasing-decreasing test, $f(x_1) < f(x_2)$ for any $a < x_1 < x_2 < b$. Therefore, $x_1 \neq x_2$ implies that $f(x_1) \neq f(x_2)$ and hence $f$ is one-to-one. So $f$ has the inverse $g = f^{-1}$.

2. The function $f$ is continuous on $(a, b)$ because it is differentiable on $(a, b)$. By the intermediate value theorem, $f$ takes all intermediate values between $f(x_1)$ and $f(x_2) > f(x_1)$ for any interval $[x_1, x_2]$ in $(a, b)$. This shows that the range of $f$ is a single interval $(c, d)$ for some $-\infty \leq c < d \leq \infty$ and if $f : [x_1, x_2] \rightarrow [f(x_1), f(x_2)]$, then $g : [f(x_1), f(x_2)] \rightarrow [x_1, x_2]$ for any $[x_1, x_2] \subset (a, b)$.

3. To show that $g$ is differentiable, fix $x \in (a, b)$ and $y = f(x) \in (c, d)$. Put $\Delta f = f(x + \Delta x) - f(x)$. Then $\Delta f/\Delta x \rightarrow f'(x)$ as $\Delta x \rightarrow 0$. Any interval $I$ with endpoints $x$ and $x + \Delta x$ is mapped by $f$ onto an interval $I'$ with the corresponding endpoints $f(x) = y$ and $f(x + \Delta x) = y + \Delta y$, where $\Delta y = \Delta f$. Put $\Delta g = g(y + \Delta y) - g(y)$. Since $g$ maps the interval $I'$ onto $I$, $\Delta g = \Delta x$. The limit $\Delta y \rightarrow 0$ implies that $\Delta x \rightarrow 0$. Therefore, $\Delta g/\Delta y = \Delta x/\Delta y = 1/(\Delta y/\Delta x) \rightarrow 1/f'(x)$ as $\Delta x \rightarrow 0$, which shows that $g'(y)$ exists, and $g'(y) = 1/f'(x)$, where $y = f(x)$. A graphic illustration is given in Figure 4.5.

For example, the exponential function $f(x) = e^x$ has the derivative $f'(x) = e^x > 0$ for all $-\infty < x < \infty$. The image of the interval $(-\infty, \infty)$ is $(0, \infty)$. So the exponential function has the differentiable inverse on $(0, \infty)$, which is, of course, the natural logarithm $\ln x$. The function $f(x) = \cos x$ has the derivative $f'(x) = -\sin x$, which is...
Figure 4.5. Inverse function theorem. An increasing function $f$, $f' > 0$, is one-to-one and hence has the inverse $g = f^{-1}$. The graphs of $f$ and $g$ are obtained from one another by the reflection about the line $y = x$. A secant line of the graph of $f$ with the slope $\tan \alpha = \Delta f / \Delta x$ is mapped on the secant line of the graph of $g$ with the slope $\tan \beta = \Delta f^{-1} / \Delta f$ by this reflection. The angles $\alpha$ and $\beta$ are related as $\alpha + \beta = \pi/2$ and hence $\tan \beta = 1 / \tan \alpha$. In the limit $\Delta x \to 0$, which also implies $\Delta f \to 0$, the secant lines become the tangent lines so that $\tan \alpha \to f'(x)$ and $\tan \beta \to g'(y)$, where $y = f(x)$. Hence, $g'(f(x)) = 1 / f'(x)$.

Negative on, for example, $(0, \pi)$. Since the image of $(0, \pi)$ is $(-1, 1)$, the inverse $\cos^{-1} x$ exists and is a differentiable function on $(-1, 1)$.

23.3. Exercises.

(1) Verify Rolle’s theorem for the function
\[ f(x) = (x - 1)(x - 2)(x - 3). \]

(2) The function $f(x) = 1 - \sqrt{x^2}$ vanishes at $x = a - 1$ and $x = b = 1$; nevertheless, $f'(x) \neq 0$ in the interval $(-1, 1)$. Does this example contradict Rolle’s theorem? Explain.
(3) Is the following assertion true? If so, prove it.
If a function \( f \) has a derivative at each point of an open interval \((a, b)\), if \( f \) is continuous at \( a \) and at \( b \), and if \( f(a) = f(b) = 0 \), then there is a point \( c \), with \( a < c < b \), such that \( f'(c) = 0 \).

(4) Is the following assertion true? If so, prove it.
If a function \( f \) has a derivative at each point of an open interval \((a, b)\) and if \( f \) is continuous at \( a \) and at \( b \), then there is a point \( c \), with \( a < c < b \), such that \( f(b) = f(a) + (b - a)f'(c) \).

(5) Let \( f \) be a function whose domain of definition is the closed interval \([a, b]\) and which is differentiable at each point of this interval. Show that if \( f'(a) > 0 \), then \( f(a) \) is a relative minimum value of \( f \).

Remark. The derivative at \( a \) is defined as the right limit,

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

(6) Let \( f \) be a function whose domain of definition is the closed interval \([a, b]\), which is continuous at \( a \) and \( b \), which is differentiable in the interval \((a, b)\), and for which \( f'(a) = 0 \) or does not exist. Show that if there is a point \( c \) such that \( f'(x) > 0 \) for \( a < x < c \), then \( f(a) \) is a relative minimum value for \( f \).

Give examples of functions with nonexisting \( f'(a) \) to illustrate this result.

(7) State and prove results similar to those given in the two previous exercises for relative maxima instead of minima and also for \( b \) instead of \( a \).

(8) Investigate whether each of the following equations has solutions. If so, for each solution, find an interval in which no other solution lies.

(i) \( \cos x = x^2 \)
(ii) \( e^x = 4 - x^4 - x^2 \)
(iii) \( \ln x = 4 - x^2 \)
(iv) \( 2\tan x = \tan^{-1} x \)
(v) \( x^6 + 3x^4 + 3x^2 + x - 7 = 0 \)

(9) Find the point(s) on the curve \( y = x^3 \) at which the tangent line is parallel to the secant line through the points \((-1, -1)\) and \((2, 8)\).

(10) Does the mean value theorem apply to the function \( f(x) = 1/x \), \( a \leq x \leq b \), if \( ab < 0 \)?

(11) Prove the generalized mean value theorem:

If \( f \) and \( g \) are functions that are differentiable in an interval \([a, b]\) and if \( g'(x) \neq 0 \) for any \( x \) in \([a, b]\), then there is a number...
23. THE MEAN VALUE THEOREM

\[ \frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}. \]

**Hint:** Consider the function \( h \) defined by
\[
h(x) = \frac{f(b) - f(a)}{g(b) - g(a)} \left( g(x) - g(a) \right) - \left( f(x) - f(a) \right).
\]

(12) Two hikers walked the same path in 3 hours. Prove that there was at least one moment of time when they had the same speed.

(13) Suppose that the function \( f(x) \) has a continuous derivative \( f'(x) \) in an interval \( (a, b) \). Given a point \( a < c < b \), is it possible to find a subinterval \( a < x_1 < c < x_2 < b \) such that
\[
\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f(c) ?
\]
Consider the example \( f(x) = x^3 \), \(-1 < x < 1\), and \( c = 0 \).

(14) Use the mean value theorem to prove the inequalities:
(i) \( |\sin a - \sin b| \leq |a - b| \)
(ii) \( pa^{p-1}(b - a) \leq b^p - a^p \leq p b^{p-1}(b - a) \), if \( 0 < a < b \) and \( p > 1 \)
(iii) \( |\tan^{-1} a - \tan^{-1} b| \leq |a - b| \)

(15) Suppose that \( f'(x) = m = \text{const for } -\infty < x < \infty \). Show that the only function that satisfies this condition is a linear function \( f(x) = b + mx \). Can one claim that the above assertion holds if the range of \( x \) is restricted to a disjoint interval? Explain.

(16) Show that if (i) the function \( f(x) \) has the second derivative \( f''(x) \) in an interval \([a, b]\) and (ii) \( f'(a) = f'(b) = 0 \), then there is a point in the interval \((a, b)\) such that
\[
|f''(c)| \geq \frac{4}{(b - a)^2} |f(a) - f(b)|.
\]

(17) Use the result of the previous exercise to show that if a car begins to move from some initial position and finishes the ride in \( t \) seconds, passing the distance \( s \) meters, then the absolute value of the car’s acceleration is no less than \( 4s/t^2 \) m/s\(^2\) at some moment of time.

(18) Find intervals in which the given function \( f \) is increasing or decreasing.
(i) \( f(x) = 2 + x - x^2 \)
(ii) \( f(x) = 3x - x^3 \)
(iii) \( f(x) = 2x^3 - 9x^2 + 12x + 1 \)
(iv) \( f(x) = \frac{2x}{1 + x^2} \)
(v) \( f(x) = \sqrt{x}/(x + 100), \ x > 0 \)
(vi) \( f(x) = \cos^2 x + 1 \)
(vii) \( f(x) = \cos(\pi/x) \)
(viii) \( f(x) = \sin x + x/2 \)
(ix) \( f(x) = e^x + 3e^{-x} + 2x \)
(x) \( f(x) = x^n e^{-x}, \ n > 0, \ x \geq 0 \)
(xi) \( f(x) = x^2 - \ln(x^2) \)
(xii) \( f(x) = x^{2-2} \)

(19) Show that the function \((1 + x^{-1})^x\) increases in the intervals \((−∞, −1)\) and \((0, ∞)\).

(20) Is the derivative of a monotonic function monotonic? Consider the example \( f(x) = x + \sin x \).

(21) Show that the function
\[
 f(x) = x + x^2 \sin \left( \frac{2}{x} \right) \quad \text{if } x \neq 0 \quad \text{and } \quad f(0) = 0
\]
is increasing at \( x = 0 \) \((f'(0) > 0)\) but is not increasing in any interval \((-a, a)\), where \(a > 0\) can be arbitrarily small. Graph the function.

(22) Suppose that \( f(x) \) and \( g(x) \) are differentiable so that \( f(a) = g(a) \) and \( f'(x) > g'(x) \) for \( x > a \). Show that \( f(x) > g(x) \) for \( x > a \).

(23) Use the result of the previous exercise to establish the inequalities:
(i) \( e^x > 1 + x, \ x \neq 0 \)
(ii) \( x - \frac{1}{2}x^2 < \ln(1 + x) < x, \ x > 0 \)
(iii) \( x - \frac{1}{6}x^3 < \sin x < x, \ x > 0 \)
(iv) \( \tan x > x + \frac{1}{3}x^3, \ 0 < x < \frac{\pi}{2} \)
Illustrate the above inequalities graphically.

(24) Find intervals in which the function \( f \) has a differentiable inverse \( f^{-1} \).
(i) \( f(x) = 2x^3 - 9x^2 + 12x + 1 \)
(ii) \( f(x) = x^2 + 16/x \)
(iii) \( f(x) = e^x + e^{-x} \)

24. The First and Second Derivative Tests

Suppose the critical points of a function \( f \) are known. If \( f \) is differentiable, then all critical points can be found by solving the equation \( f'(x) = 0 \). How can one figure out the nature of a critical point, that is,
whether it is a local maximum, local minimum, or none of the above?
It turns out that this question can be answered by studying the derivatives $f'$ and $f''$. In addition, many qualitative features of the graph $y = f(x)$ can be deduced from properties of the derivatives of $f$.

24. THE FIRST AND SECOND DERIVATIVE TESTS

24.1. The First Derivative Test. By the increasing-decreasing test, $f$ is increasing on interval if its derivative is positive, and $f$ is decreasing on an interval if its derivative is negative. Suppose $f'$ is continuous such that $f'(a) = m$ and $f'(b) = M$. Then, on the interval $[a, b]$, $f'$ must take all intermediate values between $m$ and $M$. Suppose $m < 0$ or $m > 0$ and $M < 0$, that is, the derivative changes its sign on the interval $[a, b]$, then $f'$ must vanish between $a$ and $b$. This means that $f$ has a critical point $a < c < b$, $f'(c) = 0$. More to the point, if the derivative $f'$ changes from negative to positive at $c$, then, according to the increasing-decreasing test, the function $f$ changes from increasing to decreasing at $c$, that is, $f(c - h) < f(c)$ and $f(c) > f(c + h)$ for some small positive $h$. We can then conclude that $f$ attains its local maximum at $c$. Similarly, if the derivative $f'$ changes from negative to positive at $c$, then $f$ changes from decreasing to increasing at $c$, $f(c - h) > f(c)$ and $f(c) < f(c + h)$, and hence $f$ attains its local minimum at $c$. Naturally, there is a possibility that $f'(c) = 0$ but $f'(x)$ does not change its sign at $c$. In such a situation, the increasing-decreasing test yields $f(c - h) < f(c) < f(c + h)$ or $f(c - h) > f(c) > f(c + h)$; that is, in either case the function $f$ has neither a local minimum nor a local maximum. The findings are summarized in the following theorem.

Theorem 4.8 (The First Derivative Test). Suppose that $c$ is a critical point of a continuous function $f$.

(I) If $f'$ changes from positive to negative at $c$, then $f$ has a local maximum at $c$.

(II) If $f'$ changes from negative to positive at $c$, then $f$ has a local minimum at $c$.

(III) If $f'$ does not change its sign at $c$, then $f$ has neither a local maximum nor a local minimum at $c$.

It is important to note that the very existence of $f'$ at $c$ is not required in the first derivative test. Recall the definition of a critical point ($f'(c) = 0$ or $f'(c)$ does not exist). In fact, in the preceding proof of the first derivative test, the condition $f'(c) = 0$ can be dropped because all that is needed to apply the increasing-decreasing test is the sign of the derivative $f'(x)$ for $x < c$ and $x > c$. For example,
Then \( f'(x) = -1 \) for \( x < 0 \) (the function is decreasing) and \( f'(x) = 1 \) for \( x > 0 \) (the function is increasing). Hence, \( f(x) \) has a minimum at \( x = 0 \), even though \( f' \) does not exist at \( x = 0 \). The *continuity hypothesis* is also crucial. Consider the function \( f(x) = 1/x^2 \) for \( x \neq 0 \) and \( f(0) = 0 \). Then \( f'(x) = -2/x^3 \) for \( x \neq 0 \) and \( f'(0) \) does not exist. So \( x = 0 \) is a critical point. The function is increasing for \( x < 0 \) because \( f' > 0 \), and it is decreasing for \( x > 0 \) because \( f' < 0 \). However, \( f \) has no maximum at \( x = 0 \) because \( f \) is discontinuous at \( x = 0 \). In fact, it attains its absolute minimum at \( x = 0 \). The first derivative test is summarized in the following table.

### The first derivative test

<table>
<thead>
<tr>
<th>Case</th>
<th>( \text{sign } f', x &lt; c )</th>
<th>( \text{sign } f', x &gt; c )</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>+</td>
<td>+</td>
<td>No extremum</td>
</tr>
<tr>
<td>II</td>
<td>+</td>
<td>−</td>
<td>Maximum</td>
</tr>
<tr>
<td>III</td>
<td>−</td>
<td>+</td>
<td>Minimum</td>
</tr>
<tr>
<td>IV</td>
<td>−</td>
<td>−</td>
<td>No extremum</td>
</tr>
</tbody>
</table>

There are plenty of mechanical analogies of the first derivative test. Let \( H(t) \) be the height (relative to the ground) of a stone thrown upward as a function of time \( t \). At the beginning, the stone moves upward so \( H' > 0 \) (the height is increasing). When the stone comes back to the ground, it moves downward so \( H' < 0 \) (the height is decreasing). Naturally, at some moment of time, the stone has to reach the maximal height. Analyze the motion of a pendulum (or a see-saw) from this point of view! The height would have two maxima and one minimum.

**Example 4.6 (Example 4.1 Revisited).** Find all local maxima and minima of \( f(x) = x^3 - x \) and the intervals on which the function is increasing or decreasing (the function is depicted in Figure 4.1).

**Solution:** 1. Since \( f \) is differentiable (it is a polynomial), all its critical points satisfy the equation

\[
f'(x) = 3x^2 - 1 = 3\left(x - 1/\sqrt{3}\right)\left(x + 1/\sqrt{3}\right) = 0.
\]

Hence, the critical points are \( c_1 = -1/\sqrt{3} \) and \( c_2 = 1/\sqrt{3} \).

2. For \( x < c_1 \), the product \( (x - c_1)(x - c_2) \) is positive (as the product of two negative numbers), and hence \( f' > 0 \) (\( f \) is increasing on \((-\infty, c_1)\)). For \( c_1 < x < c_2 \), the product \( (x - c_1)(x - c_2) \) is negative (as the product
of a negative and positive number), and hence \( f' < 0 \) (\( f \) is decreasing on \((c_1, c_2)\)). For \( x > c_2 \), the product \((x - c_1)(x - c_2)\) is positive (as the product of two positive numbers), and hence \( f' > 0 \) (\( f \) is increasing on \((c_2, \infty)\)).

3. The derivative changes from positive to negative at \( c_1 \). Therefore, \( f \) has a local maximum at \( c_1 \). The derivative changes from negative to positive at \( c_2 \). Therefore, \( f \) has a local minimum at \( c_2 \).  


**Definition 4.4 (Concavity).** The graph of a function \( f \) is called **concave upward** on an interval \( I \) if it lies above all of its tangent lines on \( I \). The graph is called **concave downward** on \( I \) if it lies below all of its tangent lines on \( I \).

Note that the notion of concavity implies that \( f \) is differentiable (otherwise, the tangent lines do not exist). If \( f \) is twice differentiable, then the concavity is determined by the sign of the second derivative \( f'' \). Suppose that the graph of \( f \) is concave upward on \( I \). Consider the tangent lines at two points \( c \) and \( c + h \) in \( I \):

\[
L_1(x) = f(c) + f'(c)(x - c), \quad L_2(x) = f(c + h) + f'(c + h)(x - c - h).
\]

The graph of \( f \) lies above the lines \( L_1 \) and \( L_2 \), that is, \( f(x) - L_1(x) > 0 \) and \( f(x) - L_2(x) > 0 \) for all \( x \) in \( I \). Putting \( x = c \) in the last inequality and \( x = c + h \) in the former one, we obtain

\[
\begin{align*}
  f(c) - L_2(c) &= f(c) - f(c + h) + f'(c + h)h > 0, \\
  f(c + h) - L_1(c + h) &= f(c + h) - f(c) - f'(c)h > 0
\end{align*}
\]

The sum of the right-hand sides of these inequalities is positive as the sum of two positive numbers:

\[
(4.10) \quad h[f'(c + h) - f'(c)] > 0 \quad \Rightarrow \quad \frac{f'(c + h) - f'(c)}{h} > 0,
\]

where the first inequality has been divided by a positive number \( h^2 \). Inequality (4.10) is true for any \( h \). Therefore, by taking the limit \( h \to 0 \), we can conclude that \( f''(c) > 0 \) if the graph is concave upward. Inequality (4.10) shows that \( f'(c + h) > f'(c) \) for \( h > 0 \) and \( f'(c) > f'(c + h) \) for \( h < 0 \). In other words, the derivative \( f' \), or the slope of the tangent line of the graph of \( f \), increases for the upward concavity, and hence \((f')' = f'' \) must be positive by the increasing-decreasing test. Similarly, the downward concavity implies that \( f'' \) is negative. It turns out that the converse is also true.
Theorem 4.9 (The Concavity Test). Let $f$ be twice differentiable on an interval $I$.

(I) If $f''(x) > 0$ for all $x$ in $I$, then the graph of $f$ is concave upward on $I$.

(II) If $f''(x) < 0$ for all $x$ in $I$, then the graph of $f$ is concave downward on $I$.

**Figure 4.6.** Concavity near a point $x_0$ at which $f''(x_0) = 0$. The graph is concave downward if $f''(x) < 0$ for $x > x_0$ and $x < x_0$ (left panel). Such a local behavior can be illustrated by $f(x) = c + (x - x_0)^4$, where $c$ is a constant, so that $f''(x) = 12(x - x_0)^2 \geq 0$. The graph is concave upward if $f''(x) < 0$ for $x > x_0$ and $x < x_0$ (right panel). Such a local behavior can be illustrated by $f(x) = c - (x - x_0)^4$ so that $f''(x) = -12(x - x_0)^2 \leq 0$.

How does the graph of $f$ look near a point $c$ where $f''(c) = 0$? There are four possibilities. First, $f''(c \pm h) > 0$ for some small $h > 0$. This means that the graph is concave upward to the left and right of $c$. As an example, consider $f(x) = x^4$. Second, $f''(c \pm h) < 0$. This implies that the graph is concave downward to the left and right of $c$. As an example, take $f(x) = -x^4$. These two cases are depicted in Figure 4.6. Third, $f''(c-h) > 0$ and $f''(c+h) < 0$, that is, the concavity changes from upward to downward (e.g., $f(x) = -x^3$). Fourth, $f''(c-h) < 0$ and $f''(c+h) > 0$, that is, the concavity changes from downward to upward (e.g., $f(x) = x^3$).

Definition 4.5 (Inflection Point). A point $P$ on the graph $y = f(x)$ is called an inflection point if $f$ is continuous there and the graph
changes from concave upward to concave downward or from concave downward to concave upward.

![Figure 4.7. Concavity near a point $x_0$ at which $f''(x_0) = 0$ (the case not depicted in Figure 4.6). An inflection point. The second derivative changes its sign at the inflection point $x = x_0$. The concavity of the graph of $f$ also changes at the inflection point. Such a local behavior can be illustrated by $f(x) = c + a(x - x_0) + (x - x_0)^3$, where $c$ and $a$ are constants, so that $f''(x) = 6(x - x_0)$ changes its sign at $x_0$. Note that $f'(x_0) = a$; i.e., $a$ defines the slope of the graph at $x = x_0$.

Let $c$ be a critical point of $f$. Suppose $f''$ is continuous near $c$. What can $f''(c)$ tell us about the nature of the critical number (local minimum or maximum)? There are three possibilities. First, $f''(c) > 0$. This means that $f''(x) > 0$ for all $x$ in some neighborhood of $c$ (by the continuity of $f''$). Hence, $f$ is concave upward near $c$; that is, its graph lies above the tangent line at $c$, which is a horizontal line because $f'(c) = 0$. So $f$ must have a local minimum. Similarly, the condition $f''(c) < 0$ implies that the concavity is downward near $c$ and $f$ has a local maximum. If $f''(c) = 0$, then the concavity may or may not change at $c$ as discussed earlier. The function may have a local maximum, a local minimum, or an inflection point; that is, no conclusion about the nature of the critical point can be reached.
THEOREM 4.10 (The Second Derivative Test). Suppose $f''$ is continuous near $c$.

(1) If $f'(c) = 0$ and $f''(c) > 0$, then $f$ has a local minimum at $c$.
(2) If $f'(c) = 0$ and $f''(c) < 0$, then $f$ has a local maximum at $c$.
(3) If $f'(c) = 0$ and $f''(c) = 0$, then $f$ may have a local maximum, a local minimum, or an inflection point.

**Figure 4.8.** Second derivative test. The graph of $f$ has a horizontal tangent line at a critical point, $f'(c) = 0$. If $f''(c) < 0$ and $f''(x)$ is continuous, then $f''(x) < 0$ is some open interval containing $c$. Hence, the graph is concave upward near $x = c$ and $f$ has a local maximum at $c$ (left panel). If $f''(c) > 0$, the graph is concave downward near $x = c$ and $f$ has a local minimum at $c$ (middle panel). If $f''(c) = 0$, the graph may have an inflection point when $f''$ changes its sign at $c$ (right panel), but $f''$ may not change its sign at $c$, and hence the behavior depicted in the left and middle panels is also possible in the case $f''(c) = 0$. The second derivative test is inconclusive. The function may have a local minimum, a local maximum, or an inflection. Examples are given in the captions of Figures 4.6 and 4.7 (with $a = 0$).

In Example 4.6, the function $f(x) = x^3 - x$ is shown to have two critical points: $x = \pm 1/\sqrt{3}$ as depicted in Figure 4.1. Since $f''(x) = 6x$, $f''(-1/\sqrt{3}) = -2\sqrt{3} < 0$ (a local maximum) and $f''(1/\sqrt{3}) = 2\sqrt{3} > 0$ (a local minimum). The function also has an inflection point at $x = 0$: $f''(x) = 6x < 0$ if $x < 0$ and $f''(x) = 6x > 0$ if $x > 0$. Note that an inflection point may not be a critical point! In other words, the tangent line at an inflection point can have any slope. In the example discussed, $f'(0) = -1$ (see also Figure 4.7).
24.3. Exercises.

(1) Find all critical points of the given function in its domain. Use either the first or second derivative test to determine whether there is a local maximum, a local minimum, or an inflection at each critical point.

(i) \( f(x) = (x - 1)^2 \)
(ii) \( f(x) = (x - 1)^3 \)
(iii) \( f(x) = a(x - 1)^4 \)
(iv) \( f(x) = (x - a)(x - b)(x - c), a < b < c \)
(v) \( f(x) = (x - a)(x - b)^2, a < b \)
(vi) \( f(x) = x^3 - 6x^2 + 9x - 4 \)
(vii) \( f(x) = 2x^3 - x^4 \)
(viii) \( f(x) = 8x^3 - 9x^2 + 1 \)
(ix) \( f(x) = x(x - 1)^2(x - 2)^2 \)
(x) \( f(x) = x^4 - 6x^2 + 8x + 2 \)
(xi) \( f(x) = x + \frac{1}{x} \)
(xii) \( f(x) = x^2 - 2/x \)
(xiii) \( f(x) = 2x/(1 + x^2) \)
(xiv) \( f(x) = (x^2 - 3x + 2)/(x^2 + 2x + 1) \)
(xv) \( f(x) = x\sqrt{1 - x^2} \)
(xvi) \( f(x) = x\sqrt{x - x^2} \)
(xvii) \( f(x) = \sqrt{2x - x^2} \)
(xviii) \( f(x) = \sqrt{x^2 - 1} \)
(xix) \( f(x) = \sqrt{x}\ln x \)
(xx) \( f(x) = x^{-1}\ln(x^2) \)
(xxi) \( f(x) = \sin(2x) + x \)
(xxii) \( f(x) = \ln x - x^3 \)
(xxiii) \( f(x) = \frac{x^2 - 4}{x+1} \)
(xxiv) \( f(x) = \cos x + \frac{1}{2}\cos(2x) \)
(xxv) \( f(x) = 10/(1 + \sin^2 x) \)
(xxvi) \( f(x) = e^x\sin x \)
(xxvii) \( f(x) = (e^x - e^{-x})/(e^x + e^{-x}) \)
(xxviii) \( f(x) = x^{2/3}(1 - x)^{1/3} \)
(xxix) \( f(x) = \tan^{-1}x - \frac{1}{2}\ln(1 + x^2) \)
(XXX) \( f(x) = |x|e^{-|x-1|} \)

(2) Find the absolute extreme values of the function on a given interval:

(i) \( f(x) = 2^x, [-1, 4] \)
(ii) \( f(x) = x^2 - 4x + 6, [-3, 10] \)
(iii) \( f(x) = |x^2 - 3 + 2|, [-10, 10] \)
(iv) $f(x) = x + \frac{1}{x}, [0.01, 100]$

(v) $f(x) = \sqrt{5 - 4x}, [-1, 1]$

(3) Prove the inequalities:
(i) $|3x - x^3| \leq 2$ for $|x| \leq 2$
(ii) $|a \sin x + b \cos x| \leq \sqrt{a^2 + b^2}$

(4) Find the greatest lower bound and the smallest upper bound of the function on a given interval:
(i) $f(x) = xe^{-0.01x}, (0, \infty)$
(ii) $f(x) = (1 + x^2)/(1 + x^4), (0, \infty)$

(5) Prove the inequality
\[
\frac{2}{3} \leq \frac{x^2 + 1}{x^2 + x + 1} \leq 2, \quad -\infty < x < \infty.
\]

(6) Find the constant $a$ for which the maximal deviation of the polynomial $p(x) = x^2 + a$ from 0 in the interval $-1 \leq x \leq 1$ is minimal.

(7) For which values of $a$ and $b$ will the function $h(x) = ax^2 + b/x^3$ have a horizontal tangent $(x, y) = (1, 5)$? Does $h$ have a relative maximum or minimum at $(1, 5)$?

(8) At a point $P$ in the first quadrant on the curve $y = 7 - x^2$, a tangent is drawn, meeting the coordinate axes at $A$ and $B$. Find the position of $P$ that makes the distance between $A$ and $B$ a minimum.

(9) Water flows out of a hemispherical basin through a hole at the bottom so that the volume of the water remaining at any time decreases at a rate proportional to the square root of the depth of the water remaining. Prove that the level of the water falls most slowly when the depth is two-thirds of the radius of the basin.

*Hint*: The volume of a spherical segment of one base is
\[
V = \frac{\pi}{3}(3Rh^2 - h^3),
\]
where $R$ is the radius of the sphere and $h$ is the height of the segment.

(10) A function $f$ is such that $f''$ is continuous on the interval $[a, b]$. The equation $f(x) = 0$ has three different solutions in the open interval $(a, b)$. Show that the equation $f''(x) = 0$ has at least one solution in $(a, b)$.

(11) If $f$ is a function that has a second derivative at each point of an interval $[a, b]$, show that there is a number $c$ inside this
interval such that
\[ f(b) = f(a) + f'(a)(b - a) + \frac{f''(c)}{2}(b - a)^2. \]

*Hint:* Consider the function \( h \) defined by
\[ h(x) = f(x) - f(a) - f'(a)(x - a) - k(x - a)^2, \]
where the number \( k \) is chosen such that \( h(b) = 0 \).

### 25. Taylor Polynomials and the Local Behavior of a Function

The tangent line approximation \( L(x) \) is the best linear approximation of \( f(x) \) near \( x = a \) because \( L(x) \) and \( f(x) \) have the same rate of change at \( a \). In the previous section, it was shown that the second derivative at \( a \) provides important information about the behavior of \( f(x) \) near \( a \), namely the concavity. The tangent line \( L(x) \) has no concavity as \( L''(x) = 0 \). The question arises whether there is a systematic method to improve the accuracy of the tangent line approximation to capture more essential features of the behavior of \( f(x) \) near \( a \) (i.e., the local behavior of \( f \)).

#### 25.1. Taylor Polynomials

The function \( L(x) \) is a polynomial of the first degree. Consider the second-degree polynomial
\[ T_2(x) = f(a) + f'(a)(x - a) + c_2(x - a)^2 = L(x) + c_2(x - a)^2, \]
where \( c_2 \) is an arbitrary coefficient. This polynomial has the same features as \( L(x) \), that is, \( T_2(a) = L(a) = f(a) \) and \( T'_2(a) = L'(a) = f'(a) \) because \( T'_2(x) = f'(a) + 2c_2(x - a) \). So it might provide a better approximation of \( f(x) \) than \( L(x) \) near \( a \) if the coefficient \( c_2 \) is chosen so that \( T_2(x) \) has the same concavity as \( f(x) \) near \( a \). By the concavity test, it is then reasonable to assume that \( T''_2(a) = f''(a) \), which yields \( 2c_2 = f''(a) \) or \( c_2 = f''(a)/2 \). The idea can be extended to a polynomial of degree \( n \):
\[ T_n(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + \cdots + c_n(x - a)^n, \]
where the coefficients are fixed by the conditions
\[ T_n(a) = f(a), \quad T'_n(a) = f'(a), \quad T''_n(a) = f''(a), \ldots, \quad T^{(n)}_n(a) = c_n. \]

The resulting polynomial is called the *nth-degree Taylor polynomial*:
\[ T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n. \]
25.2. Accuracy of Taylor Polynomials. The accuracy of the tangent line approximation is assessed in Theorem 3.15. Let us compare it with the accuracy of higher-degree Taylor polynomials. Consider Taylor polynomials of the exponential function $e^x$ near $x = 0$. Since $(e^x)' = e^x$ and $e^0 = 1$, the Taylor polynomials are

$$f(x) = e^x, \quad T_n(x) = 1 + x + \frac{1}{2} x^2 + \frac{1}{6} x^3 + \cdots + \frac{1}{n!} x^n.$$ 

Let us take a few values of $x$ near $x = 0$ and compare the values of the Taylor polynomials with the value of the function:

- $x = 1$: $f = 2.718$, $T_1 = 2.000$, $T_2 = 2.500$, $T_3 = 2.667$
- $x = -0.5$: $f = 0.607$, $T_1 = 0.500$, $T_2 = 0.625$, $T_3 = 0.604$
- $x = 0.25$: $f = 1.284$, $T_1 = 1.250$, $T_2 = 1.281$, $T_3 = 1.284$

Two observations can be made from this table. First, the accuracy increases with increasing the degree of the Taylor polynomial (reading the rows of the table from left to right). Second, lower-degree Taylor polynomials become more accurate as the argument gets closer to the point at which the Taylor polynomials are constructed (reading the columns of the table from top to bottom). For example, the approximation $e^x \approx T_3(x)$ is accurate up to four significant digits if $|x| \leq 1/4$. So the accuracy of the approximation $e^x \approx T_2(x)$ is determined by the difference $T_2 - T_3 = -x^3/6$, that is, by the next monomial to be added to $T_2$ to get the next Taylor polynomial. This observation is a characteristic feature of Taylor polynomials:

**Theorem 4.11.** Let $f$ be continuously differentiable $n + 1$ times on an open interval $I$ containing $a$. Let $f^{(n+1)}$ be bounded on $I$, $|f^{(n+1)}(x)| \leq M$. Then

$$|f(x) - T_n(x)| \leq \frac{M}{(n + 1)!}|x - a|^{n+1},$$

where $T_n$ is the Taylor polynomial at $a$.

Theorem 3.15 is a particular case of this theorem for $n = 1$. Inequality (4.11) is a consequence of the Taylor theorem whose proof is given in a more advanced calculus course. For example, what is the accuracy of the Taylor polynomial $T_5(x)$ near $a = 0$ for the exponential $e^x$ in the interval $[-1, 1]$? To get the upper bound on errors, one should take the maximal value of the right-hand side of (4.11) for $n = 5$ in the interval, that is, $(e^x)^{(5)} = e^x \leq M = e$, and $|x| \leq 1$, so the absolute error cannot exceed $e/6! \approx 0.0038$. 


25.3. Taylor Polynomials of Basic Functions. It is useful to make a list of a few Taylor polynomials of lower degrees for basic functions near $x = 0$. The derivation of the following relations is given as an exercise.

\begin{align*}
  e^x &\approx T_3(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 \\
  \sin x &\approx T_3(x) = T_4(x) = x - \frac{1}{6}x^3 \\
  \cos x &\approx T_3(x) = T_5(x) = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 \\
  (1 + x)^p &\approx T_3(x) = 1 + px + \frac{p(p-1)}{2}x^2 + \frac{p(p-1)(p-2)}{6}x^3 \\
  \tan x &\approx T_3(x) = T_4(x) = x - \frac{1}{3}x^3 \\
  \tan^{-1} x &\approx T_3(x) = T_4(x) = x - \frac{1}{3}x^3 \\
  \ln(1 + x) &\approx T_3(x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3
\end{align*}

Note that $\sin x$, $\tan x$, and $\tan^{-1} x$ are odd functions and their polynomial approximations cannot contain even powers of $x$, hence, $T_3(x) = T_4(x)$. Similarly, $\cos x$ is an even function and its polynomial approximation cannot contain odd powers of $x$.

25.4. Taylor Polynomials near Critical Points. Let $a$ be a critical point of $f$. Provided $f$ is enough times differentiable (see the hypotheses of Theorem 4.11), Taylor polynomials can be constructed near $a$. The linear term vanishes because $f'(a) = 0$. The second derivative test is easy to understand by looking at

\[ f(x) \approx T_2(x) = f(a) + \frac{1}{2}f''(a)(x - a)^2. \]

If $f''(a) > 0$, then near $a$ the graph of $f$ looks like a parabola concave upward (see the middle panel of Figure 4.8) and has a local minimum. If $f''(a) < 0$, then near $a$ the graph of $f$ looks like a parabola concave downward as depicted in the left panel of Figure 4.8 (a local maximum).

For example, $\cos x$ has a local maximum at $a = 0$, and it behaves near $a = 0$ as $\cos x \approx T_2(x) = 1 - x^2/2$.

The second derivative test is inconclusive if $f''(a) = 0$. In this case, $f(x)$ behaves near $a$ as

\[ f(x) \approx T_3(x) = f(a) + \frac{1}{6}f'''(a)(x - a)^3. \]

So, if $f'''(a) \neq 0$, $f$ has an inflection point at $a$ as depicted in the right panel of Figure 4.8. If $f'''(a) = 0$, one should look at

\[ f(x) \approx T_4(x) = f(a) + \frac{1}{24}f^{(4)}(a)(x - a)^4. \]

A function has a local maximum (minimum) at $a$ if $f^{(4)}(a) < 0$ ($f^{(4)}(a) > 0$) as the concavity does not change at $x = a$. This is to be compared with examples given in the caption of Figure 4.6. It is now
clear that the local behavior of \( f \) near its critical point is determined by a Taylor polynomial that has the first nonvanishing correction to \( f(a) \), provided the function is differentiable sufficiently many times.

**Example 4.7.** Investigate \( f(x) = x - \tan x \) near \( x = 0 \).

**Solution:** Find a Taylor polynomial for \( \tan x \) with two nontrivial terms. In this case, it is \( T_3: \tan x \approx T_3(x) = x + x^3/3 \) (see Section 25.3). Therefore, \( f(x) \approx x - T_3(x) = -x^3/3 \). So there is an inflection point at \( x = 0 \).

**25.5. Asymptotes.** How can the behavior of a function near \( a \) be analyzed if the function is not differentiable at \( a \), or not even defined at \( a \), or how does it behave in the asymptotic regions \( x \to \pm \infty \)?

**Definition 4.6 (Vertical Asymptotes).** The line \( x = a \) is a vertical asymptote of the graph \( y = f(x) \) if at least one of the limits

\[
\lim_{x \to a^\pm} f(x) \text{ is infinite (} \infty \text{ or } -\infty \).
\]

In other words, the function \( f(x) \) increases (decreases) unboundedly as \( x \) approaches \( a \) from either the left or the right. For example, the function

\[
(4.12) \quad f(x) = \frac{x(x^2 + 3)}{x^2 - 1} = \frac{x(x^2 + 3)}{(x-1)(x+1)}
\]

has two vertical asymptotes because the denominator vanishes at \( x = 1 \) and \( x = -1 \). When \( x \) approaches \(-1\) from the left, \( f(x) \) tends to \(-\infty\), while it tends to \( \infty \) if \(-1\) is approached from the right. Similarly, \( f(x) \to -\infty \) as \( x \to 1^- \) and \( f(x) \to \infty \) as \( x \to 1^+ \).

Suppose \( f \) has a vertical asymptote at \( a \). How does it behave near \( a \)? How “fast” does it diverge when \( x \) gets closer to \( a \)?

**Definition 4.7 (Asymptotic Behavior).** The functions \( f(x) \) and \( g(x) \) on an open interval \( x > a \) (including \( x > -\infty \)) or \( x < a \) (including \( x < \infty \)) are said to have the same asymptotic behavior at \( x = a \) if

\[
(4.13) \quad \lim_{x \to a^+} (f(x) - g(x)) = 0 \quad \text{or} \quad \lim_{x \to a^-} (f(x) - g(x)) = 0.
\]

In particular, if \( x \to \pm \infty \) and \( g(x) = mx + b \), then \( f \) is said to have a slant asymptote, and for \( m = 0 \), the slant asymptote is called a horizontal asymptote.

For a given \( f \), there are many \( g \) that have the same asymptotic behavior because one can always change \( g \) by adding \( h \) such that \( h(x) \to 0 \) as \( x \to a^\pm \). A practical problem is to find as simple a \( g \) as possible.
with the property (4.13). In other words, one looks for a simple way to estimate the values of \( f(x) \) near \( a \).

**Example 4.8.** Find the asymptotic behavior of the function (4.12) at \( x = \pm 1 \).

**Solution:** The function has to be investigated near \( x = \pm 1 \) and also when \( x \to \pm \infty \).

1. Near \( x = -1 \), the unbounded growth of \( f(x) \) is associated with the divergent factor \( 1/(x+1) \) so that \( f(x) = h(x)/(1 + x) \), where \( h(x) \) is finite near \( x = -1 \). Then \( f(x) \approx h(-1)/(x + 1) = g(x) \):

\[
f(x) = \frac{1}{x+1} \frac{x(x^2 + 3)}{x - 1} \approx \frac{2}{x + 1} = g(x).
\]

Apparently, \( \lim_{x \to -1^\pm} (f(x) - g(x)) = 0 \). The graphs of \( f(x) \) and \( g(x) = 2/(x + 1) \) are close near \( x = -1 \).

![Figure 4.9. Graph of \( f(x) \) given in (4.12) (the blue solid curve). It has a slant asymptote \( g(x) = x \) as \( x \to \pm \infty \) (the dashed line). In these asymptotic regions, \( f(x) \approx g(x) = x \). The function also has two vertical asymptotes \( x = 1 \) (the red vertical line) and \( x = -1 \) (the blue vertical line). The red solid curve is the graph of \( g(x) = 2/(x - 1) \), which shows the asymptotic behavior of \( f(x) \) near \( x = 1 \). In a neighborhood of \( x = 1 \), \( f(x) \approx g(x) \) near \( x = 1 \). The function \( f \) exhibits a similar behavior near \( x = -1 \) (not depicted here).](image-url)
2. Similarly, near $x = 1$

$$f(x) = \frac{1}{x-1} \left( x^2 + 3 \right) \approx \frac{2}{x-1} = g(x).$$

3. To find an asymptotic behavior when $x$ is large it is convenient to factor out the largest power of $x$ in the numerator and denominator:

$$f(x) = \frac{x \cdot x^2 \left( 1 + \frac{3}{x^2} \right)}{x^2 \left( 1 - \frac{1}{x^2} \right)} = x \frac{1 + \frac{3}{x^2}}{1 - \frac{1}{x^2}}$$

Since $u = 1/x^2$ is small in the asymptotic region, the factor $(1-u)^{-1} \approx 1 + u$ can be linearized. Therefore the asymptotic behavior of $f(x)$ is

$$f(x) \approx x \left( 1 + \frac{3}{x^2} \right) \left( 1 + \frac{1}{x^2} \right) \approx x \left( 1 + \frac{4}{x^2} \right) = x + \frac{4}{x}$$

where the terms $x^{-4}$ have been neglected. This shows that the graph has the slant asymptote $y = x$. Since $f(x) - x \approx 4/x$ is positive if $x > 0$ and is negative if $x < 0$, the graph approaches the slant asymptote from above as $x \to \infty$ and from below as $x \to -\infty$.

25.6. Asymptotic Behavior and Taylor Polynomials. Taylor polynomials also provide a powerful technique to investigate an asymptotic behavior of a function. This is illustrated by the following example.

**Example 4.9.** Investigate $f(x) = x^{-8/3}(1 - \cos x)$ near $x = 0$.

**Solution:** The factor $x^{-8/3}$ diverges as $x \to 0$, but $\cos x$ is smooth near $x = 0$ and can be approximated by the Taylor polynomial:

$$\cos x \approx T_4(x) = T_4 = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4,$$

$$f(x) \approx x^{-8/3}(1 - T_4(x)) = x^{-8/3} \left( \frac{1}{2}x^2 - \frac{1}{24}x^4 \right) = \frac{1}{2}x^{-2/3} - \frac{1}{24}x^{4/3}.$$ 

Therefore, for a sufficiently small $x$, $f(x) \approx \frac{1}{2}x^{-2/3} = g(x)$ because $f(x) - g(x) \approx \frac{1}{24}x^{4/3} \to 0$ as $x \to 0$. Note that the use of $T_2$ in place of $T_4$ would not be enough to establish the asymptotic behavior of $f$. □

25.7. Exercises.

1. Find Taylor polynomials of $f(x) = x^4 - x^3 + 5x^2 - 2x + 1$ at $x = 0$. What can be said about Taylor polynomials of a general polynomial function $P_n(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ at $x = 0$?

2. Decompose the polynomial $P(x) = 1 + 3x + 5x^2 - 2x^3$ into the sum of powers of the monomials $(1 + x)^n$, where $n$ is a nonnegative integer.
(3) Approximate \( f(x) = (1 + x + x^2)/(1 - x + x^2) \) by \( T_2(x) \) about \( x = 0 \). *Hint:* Approximate first \((1+u)^{-1}\) by \( T_2(u) \) about \( u = 0 \) and then use
\[
f(x) \approx (1 + x + x^2)T_2(u), \quad \text{where} \quad u = -x + x^2,
\]
retaining only \( x^n, \ n \leq 2 \), in the product.

(4) Find \( T_2(x) \) about \( x = 0 \) for \( f(x) = \sqrt{a^m + x}, \ a > 0 \).

(5) Find \( T_2(x) \) about \( x = 0 \) for \( f(x) = \sqrt{1 - 2x + x^2} \). *Hint:* Use \( T_2(u) \) for \( \sqrt{1+u} \) about \( u = 0 \) and set \( u = -2x + x^2 \).

(6) Find \( T_3(x) \) for \( f(x) = e^{2x-x^2} \) about \( x = 0 \). *Hint:* Use a suitable approximation of \( e^u \) by a Taylor polynomial where \( u = 2x-x^2 \).

(7) Find \( T_4(x) \) for \( f(x) = \sqrt{1+x^2} - \sqrt{1-x^2} \) about \( x = 0 \). *Hint:* Use a suitable Taylor polynomial to approximate \( \sqrt{1 \pm u} \), where \( u = x^2 \).

(8) Find the \( n \)th-degree Taylor polynomials of the given function at a specified point:
   (i) \( f(x) = \sin x, \ x = 0 \)
   (ii) \( f(x) = \cos x, \ x = 0 \)
   (iii) \( f(x) = \ln x, \ x = 1 \)
   (iv) \( f(x) = 1/x, \ x = 2 \)
   (v) \( f(x) = (1 + x)^p, \ p > 0, \ x = 0 \)
   (vi) \( f(x) = e^x - e^{-x}, \ x = 0 \)

(9) Estimate the absolute error of the approximation for a given interval:
   (i) \( \sin x \approx x - \frac{1}{6}x^3, \ |x| \leq \frac{1}{2} \)
   (ii) \( \tan x \approx x + \frac{1}{3}x^3, \ |x| \leq 0.1 \)
   (iii) \( \sqrt{1+x} \approx 1 + \frac{1}{2}x - \frac{1}{8}x^2, \ 0 \leq x \leq 1 \)
   *Hint:* In (i) and (ii), compare \( T_3(x) \) and \( T_4(x) \) about \( x = 0 \).

(10) Find an interval in which the approximation \( \cos x \approx 1 - \frac{1}{2}x^2 \) is accurate within the absolute error 0.0001.

(11) If \( f \) is twice continuously differentiable near \( x = 0 \) and \( f(0) = 0 \), find the local behavior of the function \( F(x) = f(|x|^p) \) near \( x = 0 \), where \( p > 0 \).

(12) Let \( f \) and \( g \) be twice differentiable at \( a \) and \( g(a) = 0 \). Find the second-degree Taylor polynomial for the function \( F(x) = f(g(x)) \) near \( x = a \). *Hint:* Use \( f(u) \approx T_2(u) = f(0) + f'(0)u + \frac{f''(0)}{2!}u^2 \), where \( u = g(x) \) and \( g(x) \) is also approximated by the corresponding \( T_2 \) near \( a \).

(13) Find the third-degree Taylor polynomial for the following functions at a specified point by using the results from the previous
exercises (i.e., by using Taylor polynomials of a suitably chosen argument):
(i) \( f(x) = \sin(x^3), \ x = 0 \)
(ii) \( f(x) = \sin(\sin x), \ x = 0 \)
(iii) \( f(x) = \tan(1 - \cos x), \ x = 0 \)

(14) Use Taylor polynomials to investigate the local behavior of a given function near a specified critical point (whether it has a local maximum, a local minimum, or an inflection):
(i) \( f(x) = \sin(x^4), \ x = 0 \)
(ii) \( f(x) = 1 - x^2/2 - \cos x, \ x = 0 \)
(iii) \( f(x) = \ln(1 + x) - x + x^2, \ x = 0 \)

(15) Use Taylor polynomials of successive degrees for \( f(x) = \ln(1 + x) \) near \( x = 0 \) to evaluate \( \ln 2 \). What degree is required to calculate \( \ln 2 \) correct within the absolute error \( 10^{-4} \)?

(16) Use Taylor polynomials to find the number correct within the given absolute error \( \varepsilon \):
(i) \( e, \ \varepsilon = 10^{-4} \)
(ii) \( \sin 1^\circ, \ \varepsilon = 10^{-4} \)
(iii) \( \cos 9^\circ, \ \varepsilon = 10^{-5} \)
(iv) \( \sqrt{5}, \ \varepsilon = 10^{-4} \)
(v) \( \log_{10} 11, \ \varepsilon = 10^{-5} \)

(17) Find vertical and slant asymptotes, if any, of a given function. Investigate the asymptotic behavior of the function near the points where it has vertical asymptotes and in the asymptotic regions \( x \to \pm \infty \).
(i) \( f(x) = x + 4/x \)
(ii) \( f(x) = x^2/(x^2 - 1) \)
(iii) \( f(x) = (x^3 - 3x^2)/(x^2 - 2x + 1) \)
(iv) \( f(x) = x^{2/3}(x^2 - 1)^{-1/3} \)
(v) \( f(x) = (\cos x - 1)/x^2 \)
(vi) \( f(x) = (x - \sin x)/x^4 \)

(18) Approximate the given function by a power function near a specified point by using Taylor polynomials:
(i) \( f(x) = x^{-2/3} \ln(1 + x), \ x = 0 \)
(ii) \( f(x) = x^{-4/3} \sin^2(2x^{2/3}), \ x = 0 \)
(iii) \( f(x) = x^{-5/3}(x - \tan(x^{1/3})), \ x = 0 \)
(iv) \( f(x) = [\sin(x - 1) - x + 1]/(x - 1)^3, \ x = 1 \)

(19) Suppose that the functions \( f \) and \( g \) are such that \( f(a) = g(a) = 0, \ f^{(k)}(a) = 0 \) for \( k = 1, 2, \ldots, n \), and \( g^{(k)}(a) = 0 \) for \( k = 1, 2, \ldots, m \), while \( f^{(n+1)}(a) \neq 0 \) and \( g^{(m+1)}(a) \neq 0 \).
Investigate the local behavior of the function \( h(x) = \frac{f(x)}{g(x)} \) near \( x = a \) if \( n = m \), if \( n > m \), and if \( m < n \).

26. L’Hôpital’s Rule

If a function \( f \) is not defined at \( a \), then its behavior near \( a \) depends on the limit of \( f \) as \( x \to a \), whether it is finite, infinite, or does not even exist. So this question is of importance when investigating a function. There is a special technique to answer it.

26.1. Indeterminate Forms \( \frac{0}{0} \) and \( \frac{\infty}{\infty} \). Consider the behavior of the following functions:

\[
\frac{e^x - 1}{x}, \quad \frac{1 - \cos x}{x^2}, \quad \frac{\tan x - x}{x^3} \quad \text{as} \quad x \to 0.
\]

Do they have a vertical asymptote at \( x = 0 \)? These functions have a common feature. They are ratios \( \frac{f}{g} \) of two functions \( f \) and \( g \) such that \( f(x) \to 0 \) and \( g(x) \to 0 \) as \( x \to 0 \). Similarly, one can make ratios where the limits of the numerator and denominator at a particular point are infinite:

\[
\frac{\ln x}{x^{-1}} \quad \text{as} \quad x \to 0^+.
\]

In general, a limit of the form

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

is called an indeterminate form of type \( \frac{0}{0} \) if both \( f(x) \to 0 \) and \( g(x) \to 0 \) as \( x \to a \); it is called an indeterminate form of type \( \frac{\infty}{\infty} \) if both \( f(x) \to \infty \) (or \(-\infty\)) and \( g(x) \to 0 \) (or \(-\infty\)). The limit itself may or may not exist. The following theorem provides a powerful method to study the indeterminate forms of these types.

**Theorem 4.12 (L’Hospital’s Rule).** Suppose \( f \) and \( g \) are differentiable and \( g'(x) \neq 0 \) on an open interval that contains \( a \) (except possibly at \( a \)). Suppose that

\[
\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0
\]

or that

\[
\lim_{x \to a} f(x) = \pm \infty \quad \text{and} \quad \lim_{x \to a} g(x) = \pm \infty.
\]

Then

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

if the limit on the right-hand side exists (or is infinite).
For the special case in which \( f(a) = g(a) = 0 \), the derivatives \( f' \) and \( g' \) are continuous, and \( g'(a) \neq 0 \), it is not difficult to see why l’Hospital’s rule (4.16) holds:

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

The first equality follows from \( f(a) = g(a) = 0 \), the second and third equalities are the consequence of the limit laws and the assumption that \( g'(a) \neq 0 \), and the last equality follows from the continuity of the derivatives. This simplified version of l’Hospital’s rule can be understood geometrically. The functions \( f \) and \( g \) can be approximated by their tangent lines at \( a \), \( f(x) \approx f'(a)(x - a) \) and \( g(x) \approx g'(a)(x - a) \), so that \( f(x)/g(x) \approx f'(a)/g'(a) \) near \( a \).

It is not so easy to prove the general version of l’Hospital’s rule (the proof is omitted here). L’Hospital’s rule is also valid for one-sided limits \( x \to a^\pm \) and for the limits at \( \pm \infty \). The hypotheses of l’Hospital’s rule must be verified for the corresponding limits.

What happens if \( f'(a) = g'(a) = 0? \) Apparently, the conditions of l’Hospital’s rule are satisfied for the derivatives \( f'(x) \) and \( g'(x) \) if \( f \) and \( g \) are twice differentiable. So l’Hospital’s rule may be applied again to the ratio \( f'(x)/g'(x) \). For functions differentiable many times, l’Hospital’s rule is easy to understand via the Taylor polynomials. Suppose that functions \( f \) and \( g \) are continuously differentiable sufficiently many times near \( a \). Then by Theorem 4.11 the following approximation holds

\[
\frac{f(x)}{g(x)} \approx \frac{f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + \cdots}{g(a) + g'(a)(x - a) + \frac{1}{2}g''(a)(x - a)^2 + \cdots}.
\]

If \( f(a) = g(a) = 0 \), then the limit of the ratio is determined by \( f'(a)/g'(a) \). If \( f(a) = g(a) = 0 \) and \( f'(a) = g'(a) = 0 \), then the limit is determined by \( f''(a)/g''(a) \) and so on.

**Example 4.10.** Investigate the indeterminate forms (4.14) and (4.15).

**Solution:** Let \( f(x) = e^x - 1 \) and \( g(x) = x \). Then \( f(0) = g(0) = 0 \) (the conditions of l’Hospital’s rule are fulfilled). Hence,

\[
\lim_{x \to 0} \frac{e^x - 1}{x} = \lim_{x \to 0} \frac{(e^x - 1)'}{(x)'} = \lim_{x \to 0} \frac{e^x}{1} = 1.
\]
2. Let \( f(x) = 1 - \cos x \) and \( g(x) = x^2 \) so that \( f(0) = g(0) = 0 \). Then \( f'(x) = \sin x \) and \( g'(x) = 2x \). Since \( f'(0) = 0 \) and \( g'(0) = 0 \), l'Hôpital's rule can be applied again:

\[
\lim_{x \to 0} \frac{1 - \cos x}{x^2} = \lim_{x \to 0} \frac{\sin x}{2x} = \lim_{x \to 0} \frac{(\sin x)'}{2(x')'} = \lim_{x \to 0} \frac{\cos x}{2} = \frac{1}{2}.
\]

3. Let \( f(x) = \tan x - x \) and \( g(x) = x^3 \) so that \( f(0) = g(0) = 0 \). The derivatives \( f'(x) = \sec^2 x - 1 \) and \( g'(x) = 3x^2 \) vanish at \( x = 0 \). L'Hôpital's rule can be used again to resolve the indeterminate form.

For complicated functions, taking higher-order derivatives might be quite an algebraic exercise. Sometimes, simple algebraic transformations of an indeterminate form in combination with basic limit laws may lead to the answer faster than a successive use of l'Hôpital's rule:

\[
\lim_{x \to 0} \frac{\tan x - x}{x^3} = \lim_{x \to 0} \frac{\sec^2 x - 1}{3x^2} = \lim_{x \to 0} \frac{1 - \cos^2 x}{3x^2} = \lim_{x \to 0} \frac{\sin^2 x}{3x^2} = \frac{1}{3} \left( \lim_{x \to 0} \frac{\sin x}{x} \right)^2 = \frac{1}{3}.
\]

The third equality follows from \( \cos x \to 1 \) as \( x \to 0 \), and therefore \( \cos^2 x \) in the denominator can be replaced by 1 in accord with the basic limit laws.

4. Let \( f(x) = \ln x \) and \( g(x) = x^{-1} \) so that \( f(x) \to -\infty \) and \( g(x) \to \infty \) as \( x \to 0^+ \). So the conditions of l'Hôpital’s rule are fulfilled. Therefore,

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

\[\square\]

26.2. Indeterminate Products 0 · ∞. Suppose that \( f(x) \to \infty \) and \( g(x) \to 0 \) as \( x \to a \). How can the indeterminate product \( f(x)g(x) \) be investigated when \( x \to a \)? It turns out the indeterminate product can be transformed into one of the indeterminate forms to which l'Hôpital’s rule is applicable:

\[
(4.17) \quad fg = \frac{f}{1/g} \left( \infty \cdot 0 \to \frac{\infty}{\infty} \right) \quad \text{or} \quad fg = \frac{g}{1/f} \left( \infty \cdot 0 \to 0 \right).
\]

The function \( x \ln x \) is an indeterminate product of the type 0 · ∞ as \( x \to 0^+ \). It can be transformed into an indeterminate form of the type \( \frac{\infty}{\infty} \) as in (4.15), which is then resolved by l'Hôpital’s rule (see Example 4.10). Note that, although either of the transformations in (4.17) may be applied with the subsequent use of l'Hôpital’s rule, the
technicalities involved might differ substantially. For instance, if the second option in (4.17) is applied to \( x \ln x = x/(1/\ln x) \), then

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

Although our goal has not been achieved, our effort has not been in vain. Since the left-hand side vanishes by Example 4.10, it follows that \( x \ln^2 x \to 0 \) as \( x \to 0^+ \). By repeating this procedure recursively, one can infer that

\[
\lim_{x \to 0^+} x(\ln x)^n = 0, \quad n = 1, 2, \ldots.
\]

26.3. Indeterminate Powers \( 0^0, \infty^0, \) and \( 1^\infty \). Several indeterminate forms arise from the limits of \([f(x)]^{g(x)}\) as \( x \to a\):

\[
0^0 \quad (f(x) \to 0, \ g(x) \to 0); \quad \infty^0 \quad (f(x) \to \infty, \ g(x) \to 0);
\]

\[
1^\infty \quad (f(x) \to 1, \ g(x) \to 0).
\]

Note \( e^0 = 1 \) if \( c \neq 0 \) and \( c \neq \infty \). Similarly, \( c^\infty = 0 \) if \( 0 \leq c < 1 \) and \( c^\infty = \infty \) if \( c > 1 \). The indeterminate powers can be transformed into an indeterminate product with the help of the identity \( y = e^{\ln y} \):

\[
\lim_{x \to a} [f(x)]^{g(x)} = \lim_{x \to a} e^{\ln([f(x)]^{g(x)})} = \lim_{x \to a} e^{g(x)\ln(f(x))} = e^{\lim_{x \to a} g(x)\ln(f(x))}.
\]

The limit of \( g(x)\ln(f(x)) \) is of type \( 0 \cdot \infty \) and can be treated by the rule (4.17). The procedure is illustrated with an example of the type \( \infty^0 \) indeterminate power:

\[
\lim_{x \to \infty} x^{1/x} = \lim_{x \to \infty} e^{\ln(x^{1/x})} = \lim_{x \to \infty} e^{\ln(x)/x} = e^{\lim_{x \to \infty} \ln(x)/x} = e^0 = 1.
\]

26.4. Indeterminate Differences \( \infty - \infty \). Suppose \( f(x) \to \infty \) and \( g(x) \to \infty \) as \( x \to a \). The limit of \( f(x) - g(x) \) as \( x \to a \) is called an indeterminate difference. The following transformations might be helpful to investigate it:

\[
f - g = f \left(1 - \frac{g}{f}\right) = \frac{1 - g/f}{1/f} \quad \text{or} \quad f - g = g \left(\frac{f}{g - 1}\right) = \frac{f/g - 1}{1/g}.
\]

If \( f(x)/g(x) \to 1 \), then the indeterminate difference is equivalent to an indeterminate form of type \( 0/0 \) and can be investigated by l'Hospital's rule. The limit of \( f/g \) is an indeterminate form of type \( \infty/\infty \) and can also be investigated by l'Hospital's rule. Suppose that \( f(x)/g(x) \to k \) as \( x \to a \), where \( k \) can be either a nonnegative number or \( k = \infty \). If \( k < 1 \), then \( f - g = g(f/g - 1) \to \infty \cdot (k - 1) = -\infty \); that is, \( g \) increases faster than \( f \) as \( x \to a \). If \( k > 1 \) or \( k = \infty \), then
\( f - g = g(f/g - 1) \to \infty \cdot (k - 1) = \infty \); that is, \( f \) increases faster than \( g \) as \( x \to a \). For example,

\[
\lim_{x \to 0^+} \left( \ln x + \frac{1}{x} \right) = \lim_{x \to 0^+} \frac{1}{x} \left( 1 + x \ln x \right) = \lim_{x \to 0^+} \frac{1}{x} (1 + 0) = \infty.
\]

If \( k = 1 \), then it is also possible that \( f - g \to c \), where \( c \) is a number. In this case, \( f \) and \( g \) increase asymptotically at the same rate: \( f' - g' \to 0 \).

If \( c = 0 \), the functions \( f \) and \( g \) have the same asymptotic behavior. For example,

\[
\lim_{x \to 0} \left( \frac{1}{\sin x} - \cot x \right) = \lim_{x \to 0} \frac{1}{\sin x} \left( 1 - \cos x \right) = \lim_{x \to 0} \frac{\sin x}{\cos x} = 0,
\]

where l’Hospital’s rule has been used in the second equality.

An alternative solution is obtained if the local behavior of the functions near \( x = 0 \) is approximated by the Taylor polynomials. Use \( T_2 \) to approximate \( \cos x \) and \( T_3 \) for \( \sin x \):

\[
\frac{1}{\sin x} - \cot x = \frac{1 - \cos x}{\sin x} \approx \frac{x^2/2}{x - x^3/6} = \frac{x/2}{1 - x^2/6} \approx \frac{x}{2},
\]

where \( x^2/6 \) is small as compared to 1 when \( x \) is close enough to 0 and can therefore be neglected in the denominator. This method is often technically easier than the use of l’Hospital’s rule.

### 26.5. Exercises.

(1) Find the limits:

(i) \( \lim_{x \to 0} \frac{\ln(1 + x)}{x} \)

(ii) \( \lim_{x \to 0} \frac{x - \sin x}{x^3} \)

(iii) \( \lim_{x \to \pi/4} \frac{\sin x - \cos x}{\cos(2x)} \)

(iv) \( \lim_{x \to 0} \frac{\sin(ax)}{\sin(bx)} \)

(v) \( \lim_{x \to 0} \frac{1 - \cos(x^2)}{x^2 \sin(x^2)} \)

(vi) \( \lim_{x \to 0} \frac{\sin^{-1}(ax)}{\sin(bx)} \)
(vii) \[ \lim_{x \to 0} \frac{\sin^{-1}(2x) - 2 \sin^{-1} x}{x^3} \]

(viii) \[ \lim_{x \to 0^+} \sin x \ln x \]

(ix) \[ \lim_{x \to 0^+} [1 \cos x] \ln x \]

(x) \[ \lim_{x \to 0} \frac{\sin x - x}{e^x - 1 - x^2/2} \]

(xi) \[ \lim_{x \to 0} \frac{\tan x - x}{e^{x^2} - 1} \]

(xii) \[ \lim_{x \to 0} x^2 \left( \frac{\sqrt{x + 1} - 1 - 1}{x} - \frac{1}{2} \right) \]

(xiii) \[ \lim_{x \to \infty} \left(1 + \frac{2}{x}\right)^x \]

(xiv) \[ \lim_{x \to \infty} (1 - e^{-x}) e^x \]

(xv) \[ \lim_{x \to \infty} (x - (\ln x)^n), \quad n > 0 \]

(xvi) \[ \lim_{x \to \infty} \left(x - \ln x\right)^{1/x} \]

(xvii) \[ \lim_{x \to 0^+} (\sin(ax))^{\sin(bx)}, \quad a > 0, \quad b > 0 \]

(xviii) \[ \lim_{x \to 0^+} \frac{\ln \sin(ax)}{\ln \sin(bx)}, \quad a > 0, \quad b > 0 \]

(xix) \[ \lim_{x \to 0} \frac{\ln \cos(ax)}{\ln \cos(bx)} \]

(xx) \[ \lim_{x \to 0} \frac{x \cot x - 1}{x^2} \]

(xxii) \[ \lim_{x \to a} \frac{a^x - x^a}{x - a}, \quad a > 0 \]

(2) Use approximations of basic functions by Taylor polynomials to find the limit:

(i) \[ \lim_{x \to 0} \frac{\cos x - e^{-x^2/4}}{x^4} \]

(ii) \[ \lim_{x \to 0} \frac{e^x \sin x - x(1 + x)}{x^3} \]

(iii) \[ \lim_{x \to \infty} x^{3/2} (\sqrt{x + 1} + \sqrt{x - 1} - 2\sqrt{x}) \]
(iv) \( \lim_{x \to 0} \frac{a^x + a^{-x} - 2}{x^2}, \ a > 0 \)

(v) \( \lim_{x \to \infty} \left[ x - x^2 \ln \left( 1 + \frac{1}{x} \right) \right] \)

(vi) \( \lim_{x \to 0} \left( \frac{1}{x} - \frac{1}{\sin x} \right) \)

(vii) \( \lim_{x \to 0} \frac{1}{x} \left( \frac{1}{x} - \cot x \right) \)

(3) Suppose that \( f(x) \) has the second derivative \( f''(x) \). Show that

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

(4) Let \( y \to 0 \) as \( x \to 0 \). Find the asymptotic behavior of \( y \), that is, the leading term \( y \approx Cx^n, \ C \neq 0 \), if

(i) \( y = \tan x - x \)
(ii) \( y = \tan(\sin x) - \sin(\tan x) \)
(iii) \( y = (1 + x)^x - 1 \)
(iv) \( y = 1 - e^{-1}(1 + x)^{1/x} \)

(5) If \( y = x - (a + b \cos x) \sin x \), find \( a \) and \( b \) such that \( y \approx Cx^5 \) as \( x \to 0 \), \( C \neq 0 \).

(6) Find \( a \) and \( b \) such that the approximation

\[
\cot x \approx \frac{1 + ax^2}{x + bx^2}
\]

is correct when the terms of order \( x^5 \) and higher can be neglected as \( x \to 0 \).

(7) Consider the function \( f(x) = e^{-1/x^2} \) if \( x \neq 0 \) and \( f(0) = 0 \). Show first that

\[
\lim_{x \to 0} x^ne^{-1/x^2} = 0, \ n > 0.
\]

Use this fact and the definition of derivatives to show that \( f^{(k)}(0) = 0 \) for all \( k \). Can Taylor polynomials be used to investigate the local behavior of the function near \( x = 0 \) and establish the nature of the critical point \( x = 0 \)?

(8) Find all critical points of the function

\[
f(x) = \begin{cases} 
  e^{-1/x}, & x > 0, \\
  0, & x = 0, \\
  -e^{1/x}, & x < 0,
\end{cases}
\]

and investigate the behavior of the function near them.
27. Analyzing the Shape of a Graph

To analyze the shape of a graph \( y = f(x) \), it is useful to have a clear idea of how the basic functions behave. For example, \( \sin x \) and \( \cos x \) are regular everywhere, bounded (e.g., \( |\sin x| \leq 1 \)), and periodic with a period of \( 2\pi \). In addition, \( \sin x \) has zeros at \( x = \pi n \), \( n = 0, \pm 1, \pm 2, \ldots \), while \( \cos x \) vanishes at \( \pi/2 + \pi n \). The function \( \sin x \) is odd, while \( \cos x \) is even. Their ratio \( \tan x = \frac{\sin x}{\cos x} \) is not defined at roots of \( \cos x \). How do \( \sin x \) behave, say, near \( x = \pi/2 \)? Since both \( \sin x \) and \( \cos x \) are smooth near \( x = \pi/2 \), the behavior of \( \tan x \) near \( \pi/2 \) can be understood with the help of Taylor polynomials. Put \( \Delta x = x - \pi/2 \) (the deviation of \( x \) from \( \pi/2 \)). Let us approximate

\[
\sin x \approx T_1(x) = 1 + (x - \frac{\pi}{2}) = 1 + \Delta x
\]

\[
\cos x \approx T_3(x) = -(x - \frac{\pi}{2}) + \frac{1}{6}(x - \frac{\pi}{2})^3 = -\Delta x + \frac{1}{6}(\Delta x)^3.
\]

Then

\[
\tan x \approx \frac{1 + \Delta x}{-\Delta x + (\Delta x)^2/6} = -\frac{1}{\Delta x} \frac{1 + \Delta x}{1 - (\Delta x)^2/6} \approx -\frac{1}{\Delta x} = -\frac{1}{x - \pi/2},
\]

where the second ratio in the product has been approximated by 1 because \( \Delta x \) is small. Since \( \tan(x + \pi) = \tan x \), this behavior repeats itself at near every root of \( \cos x \).

27.1. Growth of the Power, Exponential, and Logarithmic Functions. Let us compare the growth of the power function \( x^n \), the exponential function \( e^x \), and the logarithmic function \( \ln x \) as \( x \to \infty \). The exponential function grows faster than the power function. Let \( f(x) = e^x \) and \( g(x) = x^n \). Let us analyze the ratio \( f/g \) as \( x \to \infty \). The conditions of l’Hospital’s rule are satisfied: \( e^x \to \infty \) and \( x^n \to \infty \) as \( x \to \infty \). L’Hospital’s rule can successively be applied until the indeterminate form is resolved:

\[
\lim_{x \to \infty} \frac{e^x}{x^n} = \lim_{x \to \infty} \frac{e^x}{n x^{n-1}} = \lim_{x \to \infty} \frac{e^x}{n(n-1)x^{n-2}} = \cdots = \lim_{x \to \infty} \frac{e^x}{n!} = \infty.
\]

The conclusion is true for any real \( n \). For any real \( n \), there exists a positive integer \( N \) such that \( n < N \) or \( x^n < x^N \), \( x > 1 \). But \( e^x \) grows faster than \( x^N \). Similarly, it is straightforward to show that the logarithmic function grows slower than any power function:

\[
\lim_{x \to \infty} \frac{\ln x}{x^n} = \lim_{x \to \infty} \frac{(\ln x)'}{x^n} = \lim_{x \to \infty} \frac{\frac{1}{x}}{n x^{n-1}} = \lim_{x \to \infty} \frac{1}{n x^{n-1}} = 0
\]

for any \( n > 0 \) (\( n \) may be any positive real number here).
27.2. Asymptotes at \( x \to \pm \infty \). The asymptotic behavior of rational functions is easily determined by the highest powers of the numerator and denominator, as in Example 4.8. In general, if \( \lim_{x \to \infty} f(x) \) is infinite, then the limit of \( f/g \) can be studied for trial \( g \)s with different growth, \( g = mx \) (for slant asymptotes), \( g = x^n \), \( g = \ln x \), and so on. Suppose \( g(x) \) is found such that \( f(x)/g(x) \to 1 \) as \( x \to \infty \). Does this mean that \( g \) and \( f \) have the same asymptotic behavior? The answer is “no.” If the indeterminate form \( f(x) - g(x) \) of type \( \infty - \infty \) converges to 0 as \( x \to \infty \), then the indeterminate form \( f(x)/g(x) \) of type \( \infty/\infty \) converges to 1. Indeed, it follows from \( 1/g(x) \to 0 \) and \( f(x) - g(x) \to 0 \) that \( (1/g(x))(f(x) - g(x)) = f(x)/g(x) - 1 \to 0 \). The converse is not true. Consider the following simple example: \( f(x) = x + \sin x \) and \( g(x) = x \). Evidently, \( f(x)/g(x) = 1 + \sin x/x \to 1 \) as \( x \to \infty \). But the limit \( \lim_{x \to \infty}(f(x) - g(x)) = \lim_{x \to \infty}\sin x \) does not exist. So, even if \( g \) is found to have the property \( f(x)/g(x) \to 1 \) as \( x \to \infty \), the indeterminate form \( f - g \) of type \( \infty - \infty \) must still be investigated in order to determine whether or not \( g \) has the same asymptotic behavior as \( f \).

27.3. Guidelines for Analyzing the Shape of a Graph. The following guidelines are useful for sketching the graph of a function. It should be noted that not all the steps can always be carried out. This depends very much on the complexity of the function in question. So these are really guidelines, not a “must-do” algorithm. Given a function \( f \), find:

\[ \text{(I) Domain.} \]

The domain consists of all values of \( x \) at which \( f(x) \) is defined. Typically, it is a collection of intervals. If \( f \) is defined for \( x > a \) or \( x < a \), or both, but not at \( a \), then the local behavior of \( f \) near \( a \) must be studied (see below).

\[ \text{(II) Roots of } f \text{ and the value } f(0). \]

Roots of \( f(x) \) define the intercepts of the graph \( y = f(x) \) with the \( x \) axis. They are not always easy to find. The value \( f(0) \) (if \( x = 0 \) in the domain of \( f \)) defines the intercept of \( y = f(x) \) with the \( y \) axis.

\[ \text{(III) Symmetry and periodicity.} \]

If \( f(-x) = f(x) \) (an even function) for all \( x \) in the domain, then the graph \( y = f(x) \) is symmetric about the \( y \) axis. If \( f(-x) = -f(x) \) (an odd function) for all \( x \) in the domain, then the graph \( y = f(x) \) is symmetric about the origin (or the rotation through \( 180^\circ \) about the origin). If there is a number \( p \) such that \( f(x+p) = f(x) \), then \( f \) is periodic and \( p \) is its period.
The graph \( y = f(x) \) repeats itself on intervals of length \( p \), for example \([a, a+p] \), \([a+p, a+2p] \), and so on for any \( a \). Examples are \( \sin x, p = 2\pi \); \( \tan x, p = \pi \); \( \cos(4x), p = 2\pi/4 = \pi/2 \).

(IV) **Asymptotes and asymptotic behavior of \( f \).**

If \( f \) is a ratio \( f = h/g \), then vertical asymptotes are \( x = c \), where \( c \) solves \( g(c) = 0 \) and \( h(c) \neq 0 \). If \( h(c) = 0 \), find the limits \( \lim_{x \to c^\pm} f(x) \). If one of the limits or both is infinite, investigate the local behavior of \( f \) near \( c \) (e.g., with the help of Taylor polynomials if possible). The asymptotic behavior of \( f(x) \) near \( c \) and for large positive and negative \( x \) determines the shape of \( y = f(x) \) near the vertical asymptotes and the asymptotic shape of the graph when \( x \to \pm \infty \).

(V) **Critical points of \( f \).**

Critical points are solutions of \( f'(x) = 0 \) or the values of \( x \) where \( f'(x) \) does not exist. If, for example, \( f'(x) \) tends to \( \infty \) (or \( -\infty \)) as \( x \) approaches \( c \), then the line tangent to the graph \( y = f(x) \) at \( x = c \) is vertical. For example, \( f(x) = x^{1/3} \) and \( f'(x) = 1/(3x^{2/3}) \). So \( f'(x) \) diverges as \( x \to 0 \). The graph \( y = x^{1/3} \) has a vertical tangent line at \( x = 0 \).

(VI) **Intervals of positive and negative values of \( f \).**

These are the intervals where the graph \( y = f(x) \) lies above or below the \( x \) axis. Roots of \( f \) generally separate the intervals of positive and negative values of \( f \). However, this is not always the case. Let \( c \) be a root of \( f \). If \( f'(c) \neq 0 \), then the function \( f \) is increasing or decreasing at \( c \) and hence must change its sign. If \( f'(c) = 0 \) or \( f' \) does not exist at \( c \), that is, a root of \( f \) coincides with its critical point, then \( f \) is negative near \( c \) if \( f \) has a local maximum at \( c \) and \( f \) is positive near \( c \) if it has a local minimum at \( c \). So the sign of the derivative \( f' \) must be investigated near \( c \) (the first derivative test). Vertical asymptotes can also separate intervals of positive and negative values of \( f \). For example, the function (4.12) has one root \( x = 0 \) and two vertical asymptotes at \( x = -1 \) and \( x = 1 \). So \( f \) is negative on \((-\infty, -1)\), positive on \((-1, 0)\), negative on \((0, 1)\), and positive on \((1, \infty)\). The graph is shown in Figure 4.9.

(VII) **Intervals of increase \( (f' > 0) \) and decrease \( (f' < 0) \).**

If \( f' > 0 \) \( (f' < 0) \) on an interval, then \( f \) increases (decreases) on it (the increasing-decreasing test). These intervals are generally separated by critical points and vertical asymptotes. As a consequence of this study, the nature of each critical point is established by the first derivative test.
(VIII) **Intervals of upward and downward concavity.**
These intervals are separated by inflection points and vertical asymptotes. The sign of \( f''(x) \) must be studied. Yet, the second derivative test and Taylor polynomials can be used to establish the nature of a critical point of \( f \).

(IX) **Values of \( f \) at critical points and inflection points.**
These values set relative scales of the graph (e.g., they show how much the function increases between two critical points).

**Example 4.11.** Sketch the graph of \( f(x) = x^{1/3}(x - 6)^{2/3} \).

**Solution:** Following the preceding guidelines:
(I) The domain is the whole real line.
(II) The roots of \( f \) are \( x = 0 \) and \( x = 6 \) (the intercepts with the \( x \) axis). The intercept with the \( y \) axis is \( f(0) = 0 \).
(III) The function is not periodic, and it is neither odd nor even.
(IV) There is no vertical asymptote. To study the asymptotic behavior as \( x \to \pm \infty \), it is convenient to factor out the largest power of \( x \): \( f(x) = x(1 - 6/x)^{2/3} \) and approximate the second factor using Taylor polynomials in \( u = -6/x \). One has
\[
(1 + u)^{2/3} \approx T_2(u) = 1 + \frac{2}{3} u - \frac{1}{5} u^2 = 1 - \frac{4}{x} - \frac{4}{x^2}
\]
\[
f(x) \approx x T_2(u) = x - 4 - \frac{4}{x}
\]
This shows that the graph has a slant asymptote of the form \( y = x - 4 \). It also follows that \( f(x) - (x - 4) \approx -4/x < 0 \) if \( x \) is large and positive. Hence, the graph approaches the slant asymptote from below. Similarly, \( f(x) - (x - 4) > 0 \) if \( x \) is large and negative. Hence, the graph approaches the slant asymptote from above.
(V) The derivative reads
\[
f'(x) = \frac{x - 2}{x^{2/3}(x - 6)^{1/3}}.
\]
It vanishes at \( x = 2 \) and does not exist at \( x = 0 \) and \( x = 6 \). The critical points are 0, 2, and 6. In particular, \( f'(x) \to \infty \) as \( x \to 0 \) and it tends to \( \pm \infty \) as \( x \to 6 \), respectively. Therefore, the graph has vertical tangent lines at \( x = 0 \) and \( x = 6 \). Near \( x = 0 \), the graph looks like \( y = f(x) \approx 6^{2/3} x^{1/3} \), while near \( x = 6 \), it has a downward cusp \( y = f(x) \approx 6^{1/3} (x - 6)^{2/3} \).
(VI) The graph lies below the \( x \) axis on \( (-\infty, 0) \) as \( f(x) < 0 \) and above it on \( (0, \infty) \) as \( f(x) \geq 0 \). The function does not change
Figure 4.10. Graph of $f(x) = x^{1/3}(x-6)^{2/3}$. The roots of $f$ are $x = 0$ and $x = 6$, and they define the intercepts with the $x$ axis. It has the slant asymptote $f(x) \approx x - 4$ as $x \to \pm \infty$. The derivative vanishes at $x = 2$ (a local maximum). It diverges at $x = 0$ and $x = 6$; the graph has vertical tangent lines at these points. The second derivative is negative if $x < 0$ so that the graph is downward concave. It is positive on $(0, 6)$ and $(6, \infty)$. The graph is concave downward. The point $x = 0$ is an inflection point as the concavity changes at it.

its sign at the root $x = 6$ ($f$ must have a local minimum at 6, which is also verified by the first derivative test below).

(VII) The derivative is a product of three factors $x - 2$, $x^{-2/3}$, and $(x - 6)^{-1/3}$. By investigating the signs of these factors on the intervals separated by the critical points, we can conclude that $f' > 0$ ($f$ is increasing) on $(-\infty, 0)$, $f' > 0$ ($f$ is increasing) on $(0, 2)$, $f' < 0$ ($f$ is decreasing) on $(2, 6)$, and $f' > 0$ ($f$ is increasing) on $(6, \infty)$. Also, $f$ has a local maximum at $x = 2$ and a local minimum at $x = 6$ by the first derivative test.
The second derivative reads

$$f''(x) = -\frac{8}{x^{5/3}(x - 6)^{4/3}}.$$ 

The factor \((x - 6)^{4/3}\) cannot be negative. The sign of \(f''\) is determined only by that of \(x^{5/3}\). Thus, \(f'' > 0\) on \((-\infty, 0)\) (the graph is upward concave) and \(f'' < 0\) on \((0, 6)\) and \((6, \infty)\) (the graph is downward concave). So \(x = 0\) is the inflection point. Also, near \(x = 2\), the graph looks like the downward parabola

$$y = T_2(x) = f(2) + f''(2)(x - 2)^2 / 2 = (4 - \frac{1}{4}(x - 2)^2)/\sqrt{2}. \quad \Box$$

In the age of graphing calculators, the preceding guidelines might look rather obsolete because finding the shape of a graph can be done just by hitting the appropriate calculator buttons. But what a calculator cannot do is to provide details of the local behavior of a function near points of interest (e.g., critical points, asymptotes, etc.). In science and engineering, this is often much more important than the overall shape of a graph. In the previous example, a calculator would show that there is a slant asymptote, a cusp at \(x = 6\), and a local maximum at \(x = 2\), but it would not be able to determine the local behavior of the function near the cusp, or at the local maximum, or in the asymptotic region. Here a good working knowledge of calculus becomes indispensable, while a graphing calculator is just a useful tool that greatly facilitates the study of a function.

27.4. Exercises.

(1) Sketch the graph of each of the following functions:

(i) \(f(x) = 3x - x^3\)

(ii) \(f(x) = 1 + x^2 - \frac{x^4}{2}\)

(iii) \(f(x) = x^2(x - 1)(2 - x)\)

(iv) \(f(x) = (x + 1)(x - 2)^2\)

(v) \(f(x) = x^2/(4 + x^2)\)

(vi) \(f(x) = x^3/(1 + x^2)^2\)

(vii) \(f(x) = x/[(1 + x)(1 - x)^2]\)

(viii) \(f(x) = x^3/(x^2 - 3x + 2)\)

(ix) \(f(x) = (x^5 - x^2)/(x + 1)\)

(x) \(f(x) = [(1 + x)/(1 - x)]^4\)

(xi) \(f(x) = \sqrt{x - a}^2 + b^2\)

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

(xiii) \(f(x) = \sqrt{x^3 - x^2 - x + 1}\)

(xiv) \(f(x) = \sqrt{x^2 - \sqrt{x^2 + 1}}\)
(xv) \( f(x) = |1 + x|^{2/3}/\sqrt{x} \)
(xvi) \( f(x) = x^{-1/3}(x - 6)^{-2/3} \)
(xvii) \( f(x) = \sqrt{|x^2 - 1|} - x \)
(xviii) \( f(x) = [(x - 1)/(x + 1)]^{1/3} \)
(ix) \( f(x) = \sin x + \cos^2 x \)
(x) \( f(x) = \cos x - \frac{1}{2}\cos(2x) \)
(xi) \( f(x) = \sin x/(2 + \cos x) \)
(xii) \( f(x) = \cos x/\cos(2x) \)
(xiii) \( f(x) = x\sin x \)
(xxiv) \( f(x) = 2x - \tan x \)
(xv) \( f(x) = \sin(nx)/\sin x, \ n = 2, 3, 4 \)
(xvi) \( f(x) = x + e^{-x} \)
(xvii) \( f(x) = xe^x \)
(xviii) \( f(x) = x^{2/3}e^{-x} \)
(xix) \( f(x) = (e^x + e^{-x})\cos x \)
(xxx) \( f(x) = \sin^2 x/x^2 \)
(xxii) \( f(x) = \ln x/\sqrt{x} \)
(xxiii) \( f(x) = \ln(x + \sqrt{x^2 + 1}) \)

(2) Sketch the graph of the polynomial with \( k \) real roots:

\[ f(x) = A(x - x_1)^{n_1}(x - x_2)^{n_2}\cdots(x - x_k)^{n_k}, \]

where \( A > 0 \) and \( n_1, n_2, \ldots, n_k \) are positive integers. Investigate first the case when \( n_1 = n_2 = \cdots = n_k = 1 \), then the case when one of the powers \( n_1, n_2, \ldots, n_k \) is greater than 1 (how does the graph look when this power is odd or when it is even?). Then proceed to the general case.

(3) Let \( f \) and \( g \) be second-degree polynomials such that \( f'' > 0 \) and \( g'' > 0 \). Sketch all possible shapes of the graph \( y = f(x)/g(x) \).

28. Optimization Problems

Suppose that a quantity \( Q \) depends on some variables. The problem of optimizing \( Q \) implies finding the values of the variables at which the quantity \( Q \) attains its maximal or minimal value. The simplest optimization problem arises when \( Q \) depends on a single variable \( x \) such that \( Q \) is a function \( f(x) \). Then the optimization problem is reduced to the problem of finding extreme values of \( f(x) \). The latter problem has been analyzed in Section 22. To determine extreme values of \( f \), one has to:

(I) Find all critical points of \( f \).
(II) Investigate the nature of the critical points (local minima and local maxima). The first or second derivative tests can be used for this purpose.

(III) Calculate the values of $f$ at the endpoint of the interval $[a, b]$ (if extreme values are sought only in $[a, b]$) and compare them with values of $f$ at its local maxima and minima to determine absolute extreme values of $f$.

The following test can also be used to find absolute extreme values of a function.

**Theorem 4.13 (First Derivative Test for Absolute Extreme Values).** Suppose $c$ is a critical point of a continuous function $f$ defined on an interval.

(I) If $f'(x) > 0$ for all $x < c$ and $f'(x) < 0$ for all $x > c$, then $f(c)$ is the absolute maximum value of $f$.

(II) If $f'(x) < 0$ for all $x < c$ and $f'(x) > 0$ for all $x > c$, then $f(c)$ is the absolute minimum value of $f$.

The conclusion of the theorem is easy to understand. Consider case (I). Since $f'(x) > 0$ for all $x < c$, the function increases for all $x < c$. Since $f'(x) < 0$ for all $x > c$, the function decreases for all $x > c$. By continuity of $f$, the number $f(c)$ must be the largest value of $f$. Case (II) is proved similarly.

Recall Example 4.2. This is a typical optimization problem. Its solution is rather straightforward, provided Equation (4.4) is given. Without it, the problem of finding an optimal angle for a projectile becomes far more difficult. Its major part now involves a derivation of Equation (4.4)! This is quite typical for optimization problems. As a rule, they arise in various disciplines, and their formulation as the mathematical problem of extreme values requires a specific knowledge outside mathematics, for example, the laws of physics as in Example 4.2, chemistry, biology, economics, and so on. A typical optimization problem may be split into three basic steps:

(I) Identify a variable with respect to which a quantity $Q$ is to be optimized.

(II) Use the laws of a specific discipline to express $Q$ as a function $f$ of that variable, $Q = f(x)$.

(III) Solve the mathematical problem of extreme values of $f$.

**Example 4.12.** An aluminum can has the shape of a cylinder of radius $r$ and height $h$. Design an aluminum can of volume $V = 300 \text{ cm}^3$ to minimize the cost (or the amount) of material needed to make the can.
Solution: Following the preceding guidelines:

(I) Apparently, the least amount of material is used when the surface area of the can is minimal. So one has to minimize the surface area $S$, which depends on $r$ and $h$. But the variables $r$ and $h$ are not independent because the volume is fixed.

(II) The surface area is the sum of the areas of the side, top, and bottom of the can: $S = 2\pi rh + \pi r^2 + \pi r^2 = 2\pi rh + 2\pi r^2$. The volume is $V = \pi r^2 h$. Since the volume is fixed, the variables $r$ and $h$ are related as $h = V/(\pi r^2)$. Hence, $S$ can be written as a function of the radius $r$ only:

$$S(r) = 2\pi r \frac{V}{\pi r^2} + 2\pi r^2 = \frac{2V}{r} + 2\pi r^2.$$ 

One has to find the value of $r > 0$ at which $S(r)$ attains its absolute minimum. The corresponding value of $h$ is then found from the relation $h = V/(\pi r^2)$.

(III) The function $S(r)$ is differentiable for all $r > 0$. Therefore, all its critical points are roots of the derivative:

$$S'(r) = -\frac{2V}{r^2} + 4\pi r = 4\pi r^2 \left( r^3 - \frac{V}{2\pi} \right) = 0.$$

So the critical point is

$$r_c = \left( \frac{V}{2\pi} \right)^{1/3}.$$

Since $S'(r) < 0$ for all $0 < r < r_c$ and $S'(r) > 0$ for all $r > r_c$, the function $S(r)$ attains its absolute minimum at $r_c$ by the first derivative test for absolute extreme values. The dimensions of the can with minimal costs of material for a given volume $V$ are

$$r = \left( \frac{V}{2\pi} \right)^{1/3} \approx 3.6 \text{ cm}, \quad h = \frac{V}{\pi r_c^2} = \left( \frac{4V}{\pi} \right)^{1/3} = 2r_c \approx 7.2 \text{ cm}.$$

The analysis has shown that the height and diameter of a can of a given volume must be equal in order to minimize the cost of material (or the surface area of the can). Check out a local supermarket to see if manufacturers use this fact!

This example is further illustrated on the interactive website at http://www.math.ufl.edu/~mathguy/ufcalcbook/optimize_cylinder.html.
Remark. In the previous example, $S$ has been expressed as a function of $r$. The same conclusion could be reached if $S$ is expressed as a function of the height $h$ only, that is, when the relation $r = \sqrt{V/(\pi h)}$ is substituted into the expression for the surface area to obtain $S(h)$. The critical point of $S(h)$ can be shown to be $h_c = 2r_c$. Verify this!

A Curious Fact. The preceding problem is essential to reduce waste from plastic, glass, and aluminum containers. It can be stated more generally. What is the shape of a container that has the smallest surface area at a given volume? It can be proved by the calculus of variations that such a container must be a sphere. Even in the example of an aluminum can, the optimal dimensions appear to be as close to those of a sphere as the cylindrical geometry would allow: The height and diameter are the same. Should only spherical containers be used to “go green”? To answer this question, a far more complicated optimization problem must be studied. For example, spheres are not optimal for storage and hence for transportation; rectangular containers are far better. Storage maintenance and transportation require energy (hence carbon emissions). The production waste for containers of different shapes is different. Finally, what about consumers’ reaction to spherical Coke cans in a vending machine or spherical aluminum cans in the supermarket?

28.1. Applications to Economics. In Section 19, we introduced the cost function $C(x)$, which is the cost of producing $x$ units of a certain product. The derivative $C'(x)$ is the marginal cost. It determines the cost of increasing production from $x$ units to $x + 1$ units. Let $p(x)$ be the price per unit that a company can charge if it sells $x$ units. The function $p(x)$ is also called the price function. Naturally, it is generally expected to be a decreasing function because the price per unit usually goes down when a larger number of units is sold. The total revenue $R(x) = xp(x)$ is called the revenue function. The derivative $R'(x)$ is called the marginal revenue function. It determines the change in the revenue when the number of units sold increases from $x$ to $x + 1$. Finally, the profit function

$$P(x) = R(x) - C(x) = xp(x) - C(x)$$

determines the total profit if $x$ units are sold. Its derivative $P'(x)$ determines the change in the total profit when the number of units sold increases from $x$ to $x + 1$. The standard optimization problem here is to minimize costs and maximize revenues and profit.
Example 4.13. A small store sells jeans at a price of $80 per pair. Every week 60 units are sold. The cost to the store for 60 units is $2500, including the cost of transportation. A market survey indicates that, for each $10 rebate offered to buyers, the number of units sold will increase by 20 a week. Also, the purchase and transportation costs will go down by $2 per each weekly order increase of 5 units. How large a rebate should the store offer to maximize its profit?

Solution: 1. What is known about the price function $p(x)$? First, its value at a particular number of sold units $x = x_0 = 60$ is $p_0 = p(60) = 80$. Also, if $x$ increases by an amount of $\Delta x = 20$, the price function decreases by $\Delta p = 10$ (the rebate). Thus, the ratio $m = -\Delta p/\Delta x = -1/2$ is the rate of change of $p(x)$ (the minus sign indicates the decrease in $p(x)$). So the price function is

$$p(x) = p_0 + m(x - x_0) = 80 - \frac{1}{2}(x - 60) = 110 - \frac{1}{2}x.$$  

2. What is known about the cost function $C(x)$? First, its value at a particular number of supplied units $x = x_0 = 60$ is $C_0 = C(60) = 2500$. Also, the cost function decreases by $\Delta C = 20$ if $x$ increases by $\Delta x = 5$. So the ratio $M = -\Delta C/\Delta x = -4$ is the rate of change of $C$ or the marginal cost. Therefore,

$$C(x) = C_0 + M(x - x_0) = 2500 - 4(x - 60) = 2740 - 4x.$$  

3. One has to maximize the profit function:

$$P(x) = xp(x) - C(x) = 114x - \frac{1}{2}x^2 - 2740.$$  

Since $P'(x) = 114 - x$, the function has one critical point $x = 114$ at which $P(x)$ attains its absolute maximal value by the first derivative test for absolute extreme values.

4. If $x = 114$ units can be sold, the price per unit is $p(114) = 110 - 57 = 53$; that is, the rebate should be $p(60) - p(114) = 80 - 53 = 27$. Thus, the store should offer a rebate of $27 to maximize its profit. Note also the increase in the weekly profit: $P(60) = 2300$ whereas $P(114) = 3758$.  

Remark. In fact, the linear (tangent line) approximation has been used to get the unknown price and cost functions in the previous example. This is a benefit of market surveys: They estimate the derivatives (or trends) of the price functions. Naturally, an increase in sales leads to a decrease in the demand for that particular item. So, after a successful rebate campaign, the store would need a new market survey to estimate $p'(114)$ and get the linear approximation at $x = 114$. The price may go
up then. Similarly, the cost function is generally highly nonlinear. Its linearization near a particular \( x = x_0 \) cannot be valid for all \( x > x_0 \). Indeed, in the previous example, it vanishes at \( x = 685 \) and becomes negative after that, which cannot possibly be true.

### 28.2. Exercises.

1. Among all rectangles of a given area \( S \), find one whose perimeter is minimal.
2. Two ships move along straight lines with constant speeds \( u \) and \( v \), and the angle between the lines is \( \theta \). Find the minimal distance between the ships if at some moment of time their distances from the point of intersection of the lines were equal to \( a \) and \( b \), respectively.
3. Find all angles of a right-angled triangle of the maximal area if the sum of its cathetus and hypotenuse is constant.
4. A piece of wire 1 m long is cut into two pieces. One is bent into a square and the other into a circle. Where should cuts be made if the sum of the areas of the square and circle is to be an extreme? Which of these extremes are relative maxima and which are relative minima?
5. Show that of all triangles inscribed in a circle the equilateral triangle has the greatest area.
6. A tank has the form of a cylinder with hemispherical ends. If the volume is to be \( V \) m\(^3\), what are the dimensions for a minimum amount of material?
7. The demand for a certain article varies inversely as the cube of the selling price. If the article costs 20 cents to manufacture, find the selling price that yields the maximum profit.
8. A man is in a boat 1 mile from the nearest point, \( A \), of a straight shore. He wishes to arrive as soon as possible at a point, \( B \), 3 miles along the shore from \( A \). He can row 2 miles per hour and walk 4 miles per hour. Where should he land?
9. The stiffness of a rectangular beam varies as the product of the breadth and the cube of the depth. Find the dimensions of the stiffest beam that can be cut from a cylindrical log whose radius is \( R \).
10. A factory \( A \) is located at a (shortest) distance \( a \) miles from a railroad that goes from north to south through a town \( B \). If \( C \) is the point on the railroad at a distance \( a \) from \( A \), then \( B \) is \( b \) miles to the north of \( C \). The cost of shipping by railroad is \( p \) dollars for 1 ton per mile and the shipment by truck costs
(11) If the cost per hour for fuel required to operate a given steamer varies as the cube of its speed and is $40 per hour for a speed of 10 miles per hour, and if other expenses amount to $200 per hour, find the most economical rate to operate the steamer a distance of 500 miles.

(12) A channel of width $b$ meters joins a river of width $a$ meters at a right angle. What is the maximal length of a ship that can enter the channel from the river?

(13) A railroad company agreed to run a special train for 50 passengers at a uniform fare of $10 each. In order to secure more passengers, the company agreed to deduct 10 cents from this uniform fare for each passenger in excess of the 50 (i.e., if there were 60 passengers, the fare would be $9 each). What number of passengers would give the company the maximum gross receipt?

(14) A sheet of paper for a poster is to contain 16 square feet. The margins at the top and the bottom are to be 6 inches, and those on the sides 4 inches. What are the dimensions if the printed area is to be maximal?

(15) A taxi company charges 15 cents a mile and logs 600 passenger-miles a day. Twenty-five fewer passenger-miles a day would be logged for each cent increase in the rate per mile. What rate yields the greatest gross income?

(16) Two roads intersect at right angles, and a spring is located in an adjoining field 10 m from one road and 5 m from the other. How should a straight path just passing the spring be laid out from one road to the other so as to cut off the least amount of land? How much land is cut off?

(17) Illuminance is a measure of how much the incident light illuminates the surface. If a source of light of luminosity $k$ is positioned above a plane, the illuminance at a point on the plane is $I = k \cos \theta / r^2$, where $r$ is the distance from the source to the point and $\theta$ is the angle between the light ray from the source to the point and the normal to the plane. At what distance above the center of a round dining table of radius $R$ should a light bulb be positioned in order for the table border to have a maximal illuminance?
(18) Two light sources of luminosity $k_1$ and $k_2$ are positioned at points $A$ and $B$, respectively. Find the point on the straight line segment $AB$ that has the least illuminance if the distance between the sources is $a$ (see the definition of illuminance in the previous exercise).

(19) A pointlike source of light is positioned between two nonintersecting spheres on the line connecting the centers of the spheres. If the radii of the spheres are $R$ and $r$ ($R > r$), find the position of the light source such that the sum of illuminated areas of the spheres is maximal.

(20) A rectangular box with a square base and an open top is to be made. Find the volume of the largest box that can be made from $A \text{ cm}^2$ of material.

(21) A rectangular field containing $S \text{ m}^2$ is to be fenced off along the bank of a straight river. If no fence is needed along the river, what must be the dimensions requiring the least amount of fencing?

(22) If a stone is thrown from a cliff of height $h$ at a speed $v_0 \text{ m/s}$ and an angle $\theta$ with the horizontal line, then its trajectory is a parabola:

$$y = h + x \tan \theta - x^2 \frac{g}{2v_0^2 \cos^2 \theta},$$

where $y$ is the stone height (vertical position), $x$ is the horizontal position (all the positions are in meters), and $g = 9.8 \text{ m/s}^2$ is a constant universal for all objects near the surface of the Earth (the free-fall acceleration). Compare with Equation (4.4). At what angle is a stone to be thrown to reach the maximal range at a given speed $v_0$?

### 29. Newton’s Method

Finding roots of a function $f(x)$ is an important problem in various applications. Unfortunately, an analytic solution of the equation $f(x) = 0$ is impossible in many practical cases. For example, consider $f(x) = x - e^{-x}$. The equation $f(x) = 0$ is equivalent to $x = e^{-x}$. The graphs $y = x$ and $y = e^{-x}$ intersect at some $x$ between 0 and 1. So $f(x)$ has a root. But how can it be calculated? Here we present one of the simplest methods, known as Newton’s method. It provides a recurrence relation that allows us to compute a root of a differentiable function with any desired accuracy.
29.1. Newton's Recurrence Relation for Finding a Root. Suppose \( f(x) \) has a root near \( x_0 \). Consider the tangent line approximation of \( f \) near \( x_0 \): \( L(x) = f(x_0) + f'(x_0)(x - x_0) \). It is easy to find the root of \( L(x) \), which is denoted by \( x_1 \):

\[
L(x) = 0 \quad \implies \quad x = x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.
\]

Note that the root of \( L(x) \) exists if \( f'(x_0) \neq 0 \) (otherwise, the tangent line is horizontal and cannot have any root).

![Figure 4.11. Diagram for Newton’s method. Pick \( x_0 \) near the root of \( f \). Find the tangent line of the graph of \( f \) at \( x_0 \). Determine the intersection point \( x_1 \) of the tangent line with the \( x \) axis. Find the tangent line to the graph of \( f \) at \( x_1 \) and its intersection \( x_2 \) with the \( x \) axis. By repeating this procedure a sequence of numbers \( x_0, x_1, x_2, \ldots \) is obtained that converges to the root of \( f \), provided \( x_0 \) was chosen close enough to the root.](image)

Since \( L(x) \) is only an approximation to \( f(x) \), the number \( x_1 \) is closer to the root of \( f \) than \( x_0 \), but does not coincide with it. In other words, the value \( f(x_1) \) is closer to 0 than \( f(x_0) \): \( 0 < |f(x_1)| < |f(x_0)| \) (the absolute value is necessary if the function takes negative values). Therefore, the tangent line constructed at \( x = x_1 \), \( L(x) = f(x_1) - f'(x_1)(x - x_1) \), can be expected to approximate \( f(x) \) even better near its root because \( x_1 \) is closer to the root than \( x_0 \). The root of the new tangent line is given by the same expression as before where \( x_0 \) should be replaced by \( x_1 \): \( x_2 = x_1 - f(x_1)/f'(x_1) \). The procedure may be
recursively repeated to generate a sequence of values \( x_n \):

\[
x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \ldots,
\]

provided \( f'(x_n) \neq 0 \).

**Theorem 4.14.** If \( f \) has a single root \( r \) in an open interval and \( f'(x) \neq 0 \) on the interval, then there exists \( x_0 \) sufficiently close to \( r \) such that the sequence (4.18) converges to the root

\[
\lim_{n \to \infty} x_n = r.
\]

The convergence of \( x_n \) to \( r \) means that for all large enough \( n \) the numbers \( x_n \) lie in any small interval \( r - \varepsilon < x_n < r + \varepsilon \) for any choice of \( \varepsilon > 0 \). In practice, the root \( r \) need to be found only with some accuracy. So the sequence elements need to be calculated with a particular number of significant digits; that is \( \varepsilon = 10^{-m} \) where \( m \) is some positive integer. It is sufficient to apply Newton’s recurrence until \( x_{n+1} \) and \( x_n \) agree to all the relevant decimal places. Then \( r = x_{n+1} \) is correct to the relevant decimal places.

**Example 4.14.** Find the root of \( f(x) = x - e^{-x} \) that is correct to six decimal places.

**Solution:**

1. Determine the position of the root first. The graphs \( y = x \) and \( y = e^{-x} \) intersect only once at a point between 0 and 1. So, in any open interval containing the interval \( (0, 1) \), \( f \) has only one root.
2. Verify the condition \( f'(x) \neq 0 \): \( f'(x) = 1 + e^{-x} > 0 \) for all \( x \).
3. Pick an initial value of Newton’s sequence as close to the root as possible, e.g. \( x_0 = 0 \). Then Newton’s sequence for six decimal places is:

\[
x_0 = 0, \ x_1 = 0.5, \ x_2 = 0.566311, \ x_3 = 0.567143, \ x_4 = 0.567143.
\]

So the root \( r = 0.567143 \) is correct to six decimal places (in fact, \( f(0.567143) = -4.5 \times 10^{-7} \)).

**29.2. Pitfalls in Newton’s Method.** Unfortunately, there is no unique recipe for choosing an initial point in Newton’s sequence. The choice depends very much on the function in question. In practice, it is determined by trying different values. A few possible bad behaviors of Newton’s sequence are useful to keep in mind.
Choice of the Initial Point in Newton’s Method. A poor choice of the initial point $x_0$ can produce the value of $x_1$ that is a worse approximation to the root than $x_0$. Consider, for example, the function $f(x) = x^3 - 3x^2 + 2$ in the interval $[0, 2]$ and $f(x) = 2$ when $x < 0$ and $f(x) = -2$ when $x > 2$. This is depicted in Figure 4.12. The function is continuously differentiable because $f'(x) = 3x^2 - 6x$ approaches 0 as $x \to 0^+$ and $x \to 2^-$. The function has the root $x = 1$ and $f'(x) < 0$ in the open interval $(0, 2)$. If $0 < x_0 < 2$ is close enough to either $x = 0$ or $x = 2$, then $x_1$ would be outside the interval $(0, 2)$. Note that the actual behavior of $f(x)$ outside the interval $[0, 2]$ is not relevant for the conclusion. The essential point here is that such a situation is likely to occur when $f'(x_0)$ is close to 0.

Cycles in Newton’s Method. A poor choice of the initial point may lead to a cycle in Newton’s sequence. Take $f(x) = x^3 - 2x + 2$ and $x_0 = 0$. Since $f'(x) = 3x^2 - 2$, the next elements are $x_1 = 0 - 2/(-2) = 1$, $x_2 = 1 - 1/1 = 0 = x_0$. That is, Newton’s sequence is a cyclic sequence, which never converges. The initial point must be taken closer to the root.

Instabilities of Newton’s Method. If $f'(x) \to \pm \infty$ as $x$ approaches a root $r$ (the graph $y = f(x)$ has a vertical tangent line at the root),
Newton’s sequence may oscillate around \( r \), never converging to it, or it may diverge for any initial point. To understand this phenomenon, suppose \( f(x) \) behaves near its root \( r \) as \( f(x) \approx a(x-r)^{2\nu} \), where \( a \) is a constant and \( \nu = 1/4 \). Furthermore, if \( 0 < \nu < 1/4 \), then a Newton’s sequence does not converge. More specifically, Newton’s sequence (4.18)

\[
x_{n+1} = x_n - \frac{1}{2\nu}(x_n - r) = x_n(1 - \frac{1}{2\nu}) + \frac{1}{2\nu}r
\]

can also be written as

\[
x_{n+1} - r = x_n(1 - \frac{1}{2\nu}) + \frac{1}{2\nu}r - r = q(x_n - r),
\]

where \( q = 1 - \frac{1}{2\nu} \). Apparently, the condition \( x_n \to r \) is equivalent to \( y_n = x_n - r \to 0 \). But the sequence

\[
y_{n+1} = qy_n = q^2y_{n-1} = \cdots = q^{n+1}y_0
\]

converges to zero only if \( |q| = |1 - \frac{1}{2\nu}| < 1 \) or \( \nu > 1/4 \) unless \( y_0 = 0 \) (i.e., if the root is accurately guessed!). Recall the asymptotic behavior of the exponential function \( a^x \to \infty \) as \( x \to \infty \) if \( a > 0 \) and \( a^x \to 0 \) if \( 0 < a < 1 \). For example, for \( f(x) = x^{1/3} \) (\( \nu = 1/6 \)), Newton’s sequence diverges: \( x_{n+1} = (1-3)x_n = -2x_n \) for any choice of the
initial point \( x_0 \neq 0 \) because \(|q^n| = |(-2)^n| = 2^n\). For \( f(x) = |x|^{1/2} \ (\nu = 1/4) \), Newton’s sequence oscillates \( x_{n+1} = (1 - 2)x_n = -x_n \) (see Figure 4.13).

29.3. Understanding Money Loans. Suppose that one takes a loan of \( P \) dollars (the principal) for \( n \) months with an annual interest rate of \( I\% \). What is the monthly payment? It is calculated as follows. The interest rate per month is \( x = I/12 \). For example, an annual interest rate of 6\% means that \( I = 0.06 \) and \( x = 0.06/12 = 0.005 \). Each payment includes the payment toward the principal and the interest. Let \( F_k \) be the amount yet to be paid after \( k \) monthly payments. It is called the future value of the loan. The sequence \( F_k \) satisfies the conditions: \( F_0 = P \) and \( F_n = 0 \) (the loan and interest are paid off after \( n \) payments). Let \( A \) be the monthly payment. Then

\[
F_1 = P + Px - A, \quad F_2 = F_1 + F_1x - A, \ldots, \quad F_k = F_{k-1} + F_{k-1}x - A.
\]

Here \( F_1 \) is the future value of the loan after one payment, which is the loan \( P \) plus the monthly interest \( Px \) minus the payment \( A \). After one payment, the loan value is \( F_1 \). So, after one more payment, its value is the value \( F_1 \) plus interest \( F_1x \) minus the payment \( A \), and so on. After \( n \) payments,

\[
F_n = F_{n-1}(1 + x) - A \\
= F_{n-2}(1 + x)^2 - A[(1 + x) + 1] = \\
= F_0(1 + x)^n - A[(1 + x)^n-1 + (1 + x)^{n-2} + \cdots + (1 + x) + 1] \\
= P(1 + x)^n - A \frac{(1 + x)^n - 1}{x}.
\]

where, in the last equality, the geometric sum formula

\[
s_n = 1 + q + q^2 + \cdots + q^{n-1} = \frac{q^n - 1}{q - 1}
\]

has been used. The latter is proved by noting that the sums \( qs_n = q + q^2 + \cdots + q^{n-1} + q^n \) and \( s_n \) have the same terms except 1 in \( s_n \) and \( q^n \) in \( qs_n \) so that \( qs_n - s_n = q^n - 1 \) or \( s_n = (q^n - 1)/(q - 1) \). Since \( F_n = 0 \), the monthly payment is

\[
(4.19) \quad A = \frac{Px}{1 - (1 + x)^{-n}}.
\]

For example, a loan of $200,000 for 10 years at a fixed annual interest rate of 6\% implies 120 monthly payments of $2220.41. Indeed, in Equation (4.19), substitute \( x = 0.06/12 = 0.005 \), \( n = 120 \), and \( P = \)}
200,000, then \( A \approx 2220.41004 \). The total amount paid after 10 years is 
\[ 120 \times A = \$266,449.20. \] The interest paid is \( nA - P = \$66,449.20 \).

When selling a car, a dealer might offer a monthly payment for a few years if a customer cannot afford to pay the price in full. In this case, the loan amount \( P \) is the price of the car; the monthly payment \( A \) and its number \( n \) are known. To assess the deal, one has to figure out the interest rate before signing up. It might be the case that the loan for a higher-quality car, meaning a higher price and higher monthly payments, might have a lower interest rate, than the loan for a cheaper car (smaller monthly payments). Knowing the interest rate, one can also shop for a loan at a lower rate elsewhere (e.g., banks) to buy a car. If \( A, P, \) and \( n \) are given, then \( x \) can be found by solving Equation (4.19), which can be written in a more convenient form as

\[(4.20) \quad f(x) = Px(1 + x)^n - A(1 + x)^n + A = 0.\]

In other words, this is the root-finding problem! It can be solved by Newton’s method. The number \( x \) should be found up to five decimal places, which is sufficient our purposes.

**Example 4.15.** A dealer offers a car at a price of \( \$10,000 \). It can also be sold for payments of \( \$217.42 \) per month for 5 years. There is another car being offered at a price of \( \$15,000 \), which can also be sold for payments of \( \$311.38 \) per month for 5 years. Which loan has a lower interest rate?

**Solution:** 1. For the first car, one has to find the root of Equation (4.20) if \( A = 217.42, \ P = 10,000, \) and \( n = 5 \times 12 = 60. \) It is convenient to initiate Newton’s sequence at \( x_1 = 0.01, \) which corresponds to an annual interest rate of 12% (i.e., \( I = 0.12 \) and \( x = 0.12/12 = 0.01 \)). Up to five decimal places, Newton’s method yields \( x = 0.00917, \) which corresponds to \( I = 12x = 0.11004, \) or an annual interest rate of 11%.

2. For the second car, one has to find the root of Equation (4.20) if \( A = 311.38, \ P = 15,000, \) and \( n = 5 \times 12 = 60. \) Newton’s method, initiated again at \( x_1 = 0.01, \) yields the root \( x = 0.00750 \) (up to five decimal places). This corresponds to an annual interest rate of 9%. So the second loan has a lower interest rate. \( \square \)

It is interesting to note that the car prices differ by 50%. Similarly, the monthly payments appear in a similar proportion \( 311.38/217.42 \approx 1.43 \). The offers might look like as nearly the same deal. In fact, they are not!
29.4. Exercises.

(1) Find the number of real roots of the equation and estimate the intervals that contain a single root:
   (i) $x^3 - 6x^2 + 9x - 10 = 0$
   (ii) $x^3 - 3x^2 - 9x + h = 0$
   (iii) $3x^4 - 4x^3 - 6x^2 + 12x - 20 = 0$
   (iv) $x^5 - 5x = a$
   (v) $\ln x = kx$
   (vi) $e^x = ax^2$

(2) Under what condition does the equation $x^3 + px + q = 0$ have (i) a single real root and (ii) three real roots? Depict the corresponding regions of $(p, q)$ in a plane.

(3) Find the roots of the given equation up to five decimal places:
   (i) $\cos x = 2x$
   (ii) $e^x - e^{-x} = 1 - x^2$
   (iii) $\tan^{-1} x = x^3$
   (iv) $\tan^{-1} x = \ln x$
   (v) $\ln(1 + x^2) = 4 - x$
   (vi) $x^5 + x - 4 = 0$
   (vii) $(4 - x^2)^2 - x + 4 = 0$

(4) Newton’s method is based on the linear approximation of the function at a sample point $x_n$ to generate the next point $x_{n+1}$ of Newton’s sequence. This approximation does not take into account the concavity of the function at $x_n$. Generalize Newton’s method by using the Taylor polynomial $T_2(x)$ at $x_n$ to generate $x_{n+1}$ as a root of $T_2$. Take any of the above exercises and compare the convergence of Newton’s method with its generalization (i.e., the numbers of steps needed to obtain the root correct to the same number of decimal places, e.g., 6, 7, or 10, starting with the same initial point $x_0$).

(5) Consider a loan of $250,000$ at an annual low interest rate of $4\%$ for 15 years. Find the monthly payments. The interest rate was not fixed and is subject to change so that the monthly payments may increase up to $20\%$. How much may the annual interest rate increase (percentagewise)?

(6) A car dealer offers a car at a price of $15,000$ for 36 monthly payments of $477$. What is the interest rate?

(7) The fixed annual interest rate on a mortgage is $7\%$. For how long should one take a loan if one wants to pay in total interest no more than half of the principal? Does the maximum loan period increase or decrease with increasing or decreasing
interest rate? Does the answer change if the payments will be made every 2 weeks (i.e., 30 payments per year instead of 12)?

(8) Find the root of the equation \( \tan^{-1} x = 1 - x \) correct up to four decimal places by initiating Newton’s sequence at \( x_0 = 1 \). Investigate the dependence of the number of needed iterations to achieve this accuracy on the initial point by taking \( x_0 = n \), where \( n = 1, 2, \ldots, 10 \).

30. Antiderivatives

In many practical problems, a function is to be recovered from its derivative. For example, if the velocity is given as a function of time, \( v = v(t) \), one might want to find the position as a function of time, \( s = s(t) \), where \( s'(t) = v(t) \). What is \( s(t) \)?

**Definition 4.8.** A function \( F \) is called an antiderivative of \( f \) on an interval \( I \) if \( F'(x) = f(x) \) for all \( x \) in \( I \).

For many basic functions, it is not difficult to find the corresponding antiderivative. For example, from the rule \( (x^{n+1})' = (n+1)x^n \), it follows that if \( f(x) = x^n, n \neq -1 \), the antiderivative is \( F(x) = x^{n+1}/(n+1) \). It has also been proved that \( (\ln|x|)' = 1/x \). So the function \( F(x) = \ln|x| \) is the antiderivative of \( f(x) = 1/x \) for all \( x \neq 0 \).

30.1. Uniqueness of the Antiderivative. Suppose \( F'(x) = f(x) \) for all \( x \) in an interval \( (a, b) \). Is such an \( F(x) \) unique? This question is answered by Corollary 4.1. Indeed, let \( F(x) \) and \( G(x) \) be antiderivatives of \( f(x) \), that is, \( F'(x) = G'(x) = f(x) \) on \( (a, b) \). By Corollary 4.1, \( F \) and \( G \) may only differ by a constant: \( G(x) = F(x) + C \). Recall that Corollary 4.1 does not hold for the union of disjoint intervals. Thus, any two antiderivatives of the same function may differ at most by a constant on an interval.

**Theorem 4.15.** If \( F \) is an antiderivative of \( f \) on an interval \( I \), then the most general antiderivative of \( f \) on \( I \) is

\[
F(x) + C,
\]

where \( C \) is an arbitrary constant.

For example, the general antiderivative of the power function \( x^n \), where \( n \) is a positive integer, is

\[
F'(x) = f(x) = x^n \implies F(x) = \frac{x^{n+1}}{n+1} + C
\]
because \( x^n \) is defined on the single interval \((-\infty, \infty)\). The function \( f(x) = 1/x \) is defined on the union of disjoint intervals \((-\infty, 0)\) and \((0, \infty)\). So the general antiderivative is

\[
F'(x) = f(x) = \frac{1}{x} \implies F(x) = \begin{cases} \ln x + C_1, & x > 0 \\ \ln(-x) + C_2, & x < 0 \end{cases}
\]

where \( C_1 \) and \( C_2 \) are arbitrary constants. In what follows, the domain is always assumed to be a single interval, unless stated otherwise.

The nonuniqueness of the antiderivative is not a drawback of the concept but rather a great advantage. This is explained by the following example. The velocity of a piece of chalk thrown vertically upward with a velocity of \( v_0 \) is \( v(t) = v_0 - gt \), where \( g = 9.8 \text{ m/s}^2 \) is the acceleration of a free fall. At \( t = 0 \), the chalk has a velocity of \( v(0) = v_0 \). Then it begins to slow down (\( v(t) \) decreases because of gravity). Eventually, at \( t = v_0/g \), the chalk stops and begins to fall back. If \( h(t) \) is the height of the chalk relative to the floor, then \( h'(t) = v(t) \); that is, the height is an antiderivative of \( v(t) \). It is easy to find a particular antiderivative of \( v(t) \) using the antiderivative of the power function: \( h(t) = v_0 t - gt^2/2 \) (indeed, \( h'(t) = v_0 - gt \)). What is the physical significance of the general antiderivative \( h(t) = C + v_0 t - gt^2/2 \)? It appears as if the position of the chalk relative to the floor is not uniquely determined. In particular, \( h(0) = C \) is the height at the very moment when the chalk was thrown upward. But the chalk could be thrown upward at 1 m above the floor or 2 m above it with the very same initial velocity. So, in both cases, \( v(t) \) is the same, while the \( h(t) \) are not. In the first case, \( h(0) = 1 \), whereas in the second case, \( h(0) = 2 \). Thus, the constant \( C \) can be fixed by specifying the value of the antiderivative at a particular point.

This feature of the general antiderivative can also be visualized by plotting the graphs \( y = F(x) + C \) for different values of \( C \). All such graphs are obtained from the graph \( y = F(x) \) by rigid translations along the \( y \) axis. If one demands that the graph \( y = F(x) + C \) should pass through a particular point \((x_0, y_0)\), then \( C \) is fixed: \( y_0 = F(x_0) + C \) or \( C = y_0 - F(x_0) \). For example, find \( f(x) \) if \( f'(x) = 3x^2 \) and \( f(2) = 1 \). The general antiderivative of \( 3x^2 \) is \( f(x) = x^3 + C \). From \( f(2) = 1 \), it follows that \( f(2) = 8 + C = 1 \) or \( C = -7 \). Therefore, \( f(x) = x^3 - 7 \).

### 30.2. Linearity of the Antiderivative

Let \( F \) and \( G \) be antiderivatives of \( f \) and \( g \), respectively. Then an antiderivative of \( f + g \) is \( F + G \). An antiderivative of \( kf \), where \( k \) is an arbitrary constant, is \( kF \). These properties are easily verified. Indeed, \( (F + G)' = F' + G' = f + g \) and \( (kF)' = kF' = kf \), where the linearity of the derivative has been
used. In other words, antidifferentiation is a linear operation just like differentiation itself.

30.3. Antiderivatives of Basic Functions. An antiderivative of the power function has been found by studying the derivative of the power and logarithmic functions. The idea is useful for other basis functions. Their antiderivatives can be found by reading the table of derivatives of basic functions backward, that is, from the right to left.

### Table of antiderivatives of basic functions

<table>
<thead>
<tr>
<th>$F(x)$</th>
<th>$f(x) = F'(x)$</th>
<th>$F(x)$</th>
<th>$f(x) = F'(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sin x + C$</td>
<td>$\cos x$</td>
<td>$-\cos x + C$</td>
<td>$\sin x$</td>
</tr>
<tr>
<td>$\tan x + C$</td>
<td>$(\sec x)^2$</td>
<td>$-\cot x + C$</td>
<td>$(\csc x)^2$</td>
</tr>
<tr>
<td>$\sin^{-1} x + C$</td>
<td>$\frac{1}{\sqrt{1-x^2}}$</td>
<td>$-\cos^{-1} x + C$</td>
<td>$\frac{1}{\sqrt{1-x^2}}$</td>
</tr>
<tr>
<td>$\tan^{-1} x + C$</td>
<td>$\frac{x^n}{n+1} + C$</td>
<td>$-\cot^{-1} x + C$</td>
<td>$\frac{1}{1+x^2}$</td>
</tr>
<tr>
<td>$x^n, n \neq -1$</td>
<td>$e^x$</td>
<td>$\ln</td>
<td>x</td>
</tr>
<tr>
<td>$\frac{1}{2} \ln \left</td>
<td>\frac{1+x}{1-x} \right</td>
<td>+ C$</td>
<td>$e^x$</td>
</tr>
</tbody>
</table>

As noted earlier, here $F(x)$ is given on a single interval in the domain of $f(x)$. The table of antiderivatives of basic functions combined with the linearity of antidifferentiation is a good source of antiderivatives of more complicated functions. The chain rule \((f(g(x)))' = f'(g(x))g'(x)\) can also be used to obtain antiderivatives.

**Example 4.16.** Find the general antiderivative of $x/(x^2 + 1)$.

**Solution:** By the chain rule

\[
(ln(x^2 + 1))' = \frac{1}{x^2 + 1} \cdot (x^2 + 1)' = \frac{2x}{x^2 + 1}.
\]

Therefore $F(x) = \frac{1}{2} \ln(x^2 + 1) + C$. 

**Example 4.17.** Find the general antiderivative of $f(x) = e^{-2x} + \cos(4x) + x^2/(1 + x^2)$.

**Solution:** 1. By the linearity of the antiderivative, it is sufficient to find antiderivatives of $e^{-2x}$, $\cos(4x)$, and $x^2/(1 + x^2)$. The general antiderivative is obtained by adding a general constant to the sum of the
particular antiderivatives of the previous three functions.

2. From \((e^{-2x})' = -2e^{-2x}\), it follows that \((-e^{-2x}/2)' = e^{-2x}\). Hence, an antiderivative of \(e^{-2x}\) is \(-e^{-2x}/2\).

3. Similarly, from \((\sin(4x))' = 4\cos(4x)\), it follows that an antiderivative of \(\cos(4x)\) is \(\sin(4x)/4\).

4. The table of derivatives does not appear helpful in the case of \(x^2/(1+x^2)\). However, a simple algebraic manipulation leads to the goal:

\[
\frac{x^2}{1+x^2} = \frac{1+x^2-1}{1+x^2} = 1 - \frac{1}{1+x^2}.
\]

So its antiderivative is \(x - \tan^{-1}x\). Thus, the general antiderivative reads:

\[
F(x) = -\frac{1}{2}e^{-2x} + \frac{1}{4}\sin(4x) + x - \tan^{-1}x + C.
\]

because \(f(x)\) is defined on the single interval \((-\infty, \infty)\). \(\square\)

30.4. Antiderivatives of Higher Order. What is \(F(x)\) if \(F''(x) = f(x)\) for a given \(f(x)\)? Or, more generally, what is \(F(x)\) if \(F^{(n)}(x) = f(x)\)? A function \(F\) that satisfies the latter condition is called an antiderivative of \(f\) of the \(n\)th order. To find it, one has to antidifferentiate \(f\) \(n\) times. For example, \(F''(x) = 6x\). Taking the first antiderivative of \(f(x) = 6x\), one gets \(F'(x) = 3x^2\). Taking the antiderivative one more time yields \(F(x) = x^3\). What about the uniqueness of higher-order antiderivatives? To find the general antiderivative of a higher order, each time antidiffrentiation is carried out, the corresponding general antiderivative must be used. In the preceding example, the general antiderivative of \(f(x) = 6x\) is \(3x^2 + C_1\), where \(C_1\) is an arbitrary constant. Hence, \(F'(x) = 3x^2 + C_1\). Its general antiderivative reads \(F(x) = x^3 + C_1x + C_2\), where \(C_2\) is another arbitrary constant. Thus, the general second antiderivative can be obtained from a particular one by adding a general function whose second derivative is 0, which is a general linear function: \((C_1x + C_2)' = 0\). Similarly, if \(F(x)\) is a particular function that satisfies the condition \(F^{(n)}(x) = f(x)\), then the general antiderivative of the \(n\)th order is

\[
F(x) + C_1x^{n-1} + C_2x^{n-2} + \cdots + C_{n-1}x + C_n,
\]

where \(C_1, \ldots, C_n\) are arbitrary constants. Indeed, the \(n\)th derivative of a polynomial of degree \(n-1\) is 0. Note that this analysis is justified only when \(f\) was defined in an interval. The reader is instructed to analyze the situation when the domain of the function \(f\) consists of disjoint
intervals and, in particular, to consider higher order antiderivatives of the inverse power function $x^{-n}$, where $n$ is a positive integer.

The following example illustrates the significance of arbitrary constants in general higher-order antiderivatives.

**Example 4.18.** Any free-falling object near the surface of the Earth has the free-fall acceleration of 9.8 m/s$^2$. A piece of chalk is thrown vertically upward at a speed of 7 m/s and at 1.5 m above the floor. When does the chalk hit the floor?

**Solution:** 1. Let $h(t)$ be the height of the chalk relative to the floor. Then its velocity is $v(t) = h'(t)$, and its acceleration is $a(t) = v'(t) = h''(t)$. Since all free-falling objects have an acceleration of 9.8 m/s$^2$, one has $h''(t) = -9.8$. The minus sign indicates that the acceleration is directed downward.

2. The general second antiderivative of the constant function $-9.8$ is

$$h''(t) = -9.8 \implies h(t) = -9.8 \frac{t^2}{2} + C_1 t + C_2,$$

where $C_1$ and $C_2$ are arbitrary constants.

3. To fix $C_1$ and $C_2$, the initial conditions of the motion must be used. The initial velocity is $v(0) = 7$. Since $v(t) = h'(t) = -9.8t + C_1$, one infers that $v(0) = C_1 = 7$. The initial height is $h(0) = 1.5$. Hence, $h(0) = C_1 = 1.5$.

4. The height is $h(t) = -9.8t^2/2 + 7t + 1.5$. The chalk hits the floor when its height vanishes, that is, at the time moment $t > 0$ when $h(t) = 0$. A positive root of the quadratic equation $-9.8t^2/2 + 7t + 1.5 = 0$ is $t \approx 1.62$ s. The maximum height reached by the chalk is 4 m. Why? □

### 30.5. Exercises.

(1) Use the table of antiderivatives of basic functions to find an antiderivative of each of the following functions:

(i) $f(x) = (3 - x^2)^2$
(ii) $f(x) = x^2(1 - x)^2$
(iii) $f(x) = (1 - x)(1 - 2x)(1 - 3x)$
(iv) $f(x) = \sin(4x) + x$
(v) $f(x) = 1/(x^2 + 4)$
(vi) $f(x) = e^{3x} + e^{-3x} + \sqrt{x}$
(vii) $f(x) = \cos^2 x$ (Hint: $2\cos^2 x = 1 + \cos(2x)$)
(viii) $f(x) = \sin(ax) \cos(bx)$
(ix) $f(x) = \cos(ax) \cos(bx)$
(x) $f(x) = \sin(ax) \sin(bx)$
4. APPLICATIONS OF DIFFERENTIATION

*Hint*: Express the products of trigonometric functions via the sum of trigonometric functions.

(i) \( f(x) = \frac{x + 1}{\sqrt{x}} \)
(ii) \( f(x) = \sqrt{x^2(1 + \sqrt{x})} \)
(iii) \( f(x) = \frac{x}{x^2 + 1} \)
(iv) \( f(x) = \frac{1}{(x^2 + a^2)} \)
(v) \( f(x) = \frac{x^2 + a^2}{(x^2 + b^2)} \)
(vi) \( f(x) = \frac{x^4 + 1}{(x^2 + 1)} \)

*Hint*: \( x^4 = x^4 + 2x^2 - 2x^2 \)

(vii) \( f(x) = |x| \)
(viii) \( f(x) = xe^{-x^2} \)
(ix) \( f(x) = |x - a| + |x - b|, \) where \( a < b \)
(x) \( f(x) = \sin(2x) \)
(xi) \( f(x) = 1/(x^2 - 1), \) where \(-1 < x < 1\)

*Hint*: \( 1/[(x - a)(x - b)] = A[1/(x - a) - 1/(x - b)]; \) find \( A. \)
(xii) \( f(x) = \frac{1}{\sqrt{x^2 - a^2}}, \) where \( x > a > 0 \)
(xiii) \( f(x) = 2^x/3^x \)
(xiv) \( f(x) = (2^x + 3^{-x})^2 \)

(2) Use the table of antiderivatives to find the general second antiderivative of each of the following functions:

(i) \( f(x) = x(x - 1) \)
(ii) \( f(x) = x^{1/3} \)
(iii) \( f(x) = \sin(2x) \)
(iv) \( f(x) = |x| \)
(v) \( f(x) = f_0 = \text{const if } x \in [a, b] \) and \( f(x) = 0 \) otherwise
(vi) \( f(x) = |x| + 1 \)
(vii) \( f(x) = \sin x \cos(2x) \)
(viii) \( f(x) = e^{3x} + e^{-3x} \)
(ix) \( f(x) = x/(x^2 + 1)^2 \)
(x) \( f(x) = 2^x - 3^{-x} \)

(3) Find the general antiderivative of the function whose domain is not a single interval:

(i) \( f(x) = 1/(x - 1) \)
(ii) \( f(x) = 1/(x^2 - 1) \)
(iii) \( f(x) = x^{-4/3} \)
(iv) \( f(x) = (1 - x)^3/x^{4/3} \)
(v) \( f(x) = \sec^2 x \)
(vi) \( f(x) = \csc^2 x \)

(4) A car that was at rest accelerates at a rate of 1 m/s\(^2\) for 1 minute. Then it decelerates at a rate of 0.5 m/s\(^2\) until it stops. Find the distance traveled by the car.
CHAPTER 5
Integration

31. Areas and Distances

31.1. Limit of a Numerical Sequence. A numerical sequence \( a_1, a_2, a_3, \ldots \) is a function whose domain is the set of all positive integers. In other words, it is a rule that assigns a unique number \( a_n \) to each integer \( n \). For example,

\[ a_n = \frac{n}{n + 1}, \quad n = 1, 2, 3, \ldots \]

If \( f(x) = x/(x + 1) \), then the sequence \( a_n \) is formed by the values of \( f(x) \) at \( x = n \), \( a_n = f(n) \). It follows from Section 10 that

\[ \lim_{x \to \infty} f(x) = \lim_{x \to \infty} \frac{x}{x + 1} = \lim_{x \to \infty} \frac{1}{1 + \frac{1}{x}} = 1. \]

This means that for any given \( \varepsilon > 0 \), there exists a real number \( N \) such that values of \( f(x) \) lies in the interval \( 1 - \varepsilon < f(x) < 1 + \varepsilon \) for all \( x > N \). Since \( \varepsilon \) is arbitrarily, one can say that the values of \( f \) are arbitrarily close to 1. In particular, for \( n \) large enough, the values \( a_n = f(n) \) are arbitrarily close to 1, and one can conclude that

\[ \lim_{n \to \infty} a_n = \lim_{n \to \infty} \frac{n}{n + 1} = 1. \]

Since numerical sequences are just a special case of a function, a formal definition of the limit of a sequence can be adopted from Section 10.

**Definition 5.1 (Limit of a numerical sequence).** Let \( a_n, \ n = 1, 2, \ldots, \) be a numerical sequence. The sequence is said to converge to a number \( a \) if for any \( \varepsilon > 0 \) there is an integer \( N \) such that \( |a_n - a| < \varepsilon \) for all \( n > N \), and in this case one writes

\[ \lim_{n \to 0} a_n = a \]

If the sequence has no limit, it is called divergent.

This definition also means that a convergent sequence has all but a finite number of its terms in an arbitrarily small interval \((a - \varepsilon, a + \varepsilon)\). For example, take \( a_n = 1/n^p \) where \( p > 0 \). Then evidently \( \lim_{n \to \infty} a_n = 0 \). Indeed, given any small number \( \varepsilon > 0 \), it follows that the inequality

\[ |a_n - 0| = a_n = \frac{1}{n^p} < \varepsilon \implies n > \frac{1}{\varepsilon^{1/p}} \]
holds for all \( n > N \) where \( N \) is an integer such that \( N > 1/\varepsilon^{1/p} \).
Therefore, only a finite number of terms of the sequence lie outside the interval \((-\varepsilon, \varepsilon)\) and, hence, the sequence converges to 0. This sequence is formed by the values of the function \( f(x) = 1/x^p \), \( a_n = f(n) \). In Section 10, it has been shown that \( f(x) \to 0 \) as \( x \to \infty \), which also implies that \( f(n) \to 0 \) as \( n \to \infty \). The relation between the limits of sequences and the limits of functions at infinity
\[
\lim_{x \to \infty} f(x) = a \implies \lim_{n \to \infty} f(n) = \lim_{n \to \infty} a_n = a
\]
is a useful tool to calculate the limits of sequences. Note that the converse of the above assertion is not true. Why?

A final remark is that the basic limit laws and the squeeze principle hold for the limits of sequences.

31.2. Area Under a Graph. Consider the linear function \( f(x) = x \).
What is the area below the graph \( y = f(x) \) and above the interval \( 0 \leq x \leq 1 \)? This question is easy to answer because the area in question is the area of the right triangle with catheti of unit length: \( A = 1/2 \). Let \( f(x) = x^2 \). What is the area now? To calculate it, consider a partition of the interval \([0, 1]\) by \( n \) segments on length \( 1/n \).
The partition is defined by the set of points \( x_0 = 0, x_1 = 1/n, x_2 = 2/n, \ldots, x_{n-1} = (n-1)/n, \) and \( x_n = n/n = 1 \), that is, \( x_k = k/n \), where \( k = 0, 1, 2, \ldots, n \). The area under the parabola \( y = x^2 \) over the interval \([0, 1]\) is the sum of the areas \( S_k \) under the parabola over the partition interval \([x_{k-1}, x_k]\) where \( k = 1, 2, \ldots, n \),
\[
A = S_1 + S_2 + \cdots + S_n.
\]
In the interval \([x_{k-1}, x_k]\), the function \( f(x) = x^2 \) attains its maximum value at \( x = x_k \) and its minimum value at \( x = x_{k-1} \). Therefore the area \( S_k \) cannot exceed the area of a rectangle with base \( 1/n \) and height \( f(x_k) = (k/n)^2 \). Let us denote this upper bound by \( S_k^U = k^2/n^3 \). The area \( S_k \) is greater than the area of a rectangle with base \( 1/n \) and height \( f(x_{k-1}) = (k-1)^2/n^2 \). The lower bound is denoted by \( S_k^L = (k-1)^2/n^3 \). Thus,
\[
S_k^L = \frac{(k-1)^2}{n^3} < S_k < \frac{k^2}{n^3} = S_k^U.
\]
So, the area \( A \) is bounded above by the sum of \( S_k^U \) and below by the sum of \( S_k^L \):
\[
S_1^L + S_2^L + \cdots + S_n^L = A_n^L \leq A \leq A_n^U = S_1^U + S_2^U + \cdots + S_n^U
\]
for any number \( n \) of partition segments.
Figure 5.1. The upper and lower bounds for the area under the graph \( y = f(x) = x^2 \) for \( n = 4 \) partition intervals in \([0, 1]\). The upper bound is obtained by taking the maximum value of \( f \) on each partition interval (left panel). The lower bound is obtained by taking the minimum value of \( f \) on each partition segment (right panel). When \( n \) increases, the upper bound decreases, while the lower bound decreases, both approaching the area under the graph as \( n \to \infty \).

Let us calculate the difference

\[
0 < S^U_k - S^L_k = \frac{2k - 1}{n^3} \leq \frac{2n - 1}{n^3} < \frac{2}{n^2}
\]

for any \( k = 1, 2, \ldots, n \); in the second inequality, the condition \( k \leq n \) has been used. This inequality allows us to estimate the difference \( A^U_n - A^L_n \):

\[
0 < A^U_n - A^L_n = (S^U_1 - S^L_1) + (S^U_2 - S^L_2) + \cdots + (S^U_n - S^L_n)
< \frac{2}{n^2} = \frac{2}{n}.
\]

Thus, if the limit \( \lim_{n \to \infty} A^U_n \) exists, then \( \lim_{n \to \infty} A^U_n = \lim_{n \to \infty} A^L_n \) because \( 0 < A^U_n - A^L_n < 2/n \to 0 \) as \( n \to \infty \). On the other hand, \( A^L_n \leq A \leq A^U_n \) for any \( n \). Taking the limit \( n \to \infty \) in this inequality yields

\[
\lim_{n \to \infty} A^L_n = A = \lim_{n \to \infty} A^U_n.
\]

From a geometrical point of view, when \( n \) gets larger, the area \( A^U_n \) approaches \( A \) from above while \( A^L_n \) does so from below. For \( n \) large enough, both \( A^U_n \) and \( A^L_n \) may serve as a good approximation of \( A \). In fact, the error of either of the approximations does not exceed \( 2/n \) because \( 0 < A^U_n - A^L_n < 2/n \) and \( A^L_n \leq A \leq A^U_n \). It appears that the limit \( \lim_{n \to \infty} A^U_n \) can actually be calculated by means of the formula
for the sum of squares of the first $n$ positive integers:

\begin{equation}
1^2 + 2^2 + \cdots + n^2 = \frac{1}{6} n(n + 1) (2n + 1) = \frac{n}{6} (2n^2 + 3n + 1).
\end{equation}

Indeed, by making use of this formula, one can infer that

\[
\lim_{n \to \infty} A_n^U = \lim_{n \to \infty} \frac{1}{n^3} (1^2 + 2^2 + \cdots + n^2) = \lim_{n \to \infty} \frac{2n^2 + 3n + 1}{6n^2} = \lim_{n \to \infty} \left( \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \right) = \frac{1}{3} + \lim_{n \to \infty} \frac{1}{2n} + \lim_{n \to \infty} \frac{1}{6n^2} = \frac{1}{3}
\]

So the area is $A = \frac{1}{3}$.

Let $x_k^*$ be a number in the interval $[x_{k-1}, x_k]$. Then the area $S_k$ can also be approximated by the area $S_k^*$ of a rectangle with base $1/n$ and height $f(x_k^*) = (x_k^*)^2$, that is, $S_k^* = f(x_k^*)/n$. Then the total area under the graph is approximated by the sum $A_n^*$ of all $S_k$. Since $S_k^* \leq S_k \leq S_k^U$ (owing to the monotonicity of the function $x^2$ in each interval $[x_{k-1}, x_k]$), the following inequality holds for any $n$:

\[
A_n^L \leq S_1^* + S_2^* + \cdots + S_n^* = A_n^* \leq A_n^U.
\]

Taking the limit $n \to \infty$ in this inequality and using the squeeze principle, a remarkable result is obtained

\[
\lim_{n \to \infty} A_n^* = A;
\]

that is, the limit of $A_n^*$ does not depend on the choice of sample points $x_k^*$. The area could have been approximated by, for example, $A_n^*$ with the sample points as the midpoints $x_k^* = (x_k + x_{k-1})/2$, or any other convenient choice. This analysis can be extended to any continuous function.

The calculation of the area under the graph is further illustrated in the video website at http://www.math.ufl.edu/~mathguy/ufcalcbook/riemann.html.

31.3. The Area Under the Graph of a Continuous Function. Let $f(x)$ be continuous on $[a, b]$. Consider a partition of $[a, b]$ by $n$ segments of length $\Delta x = (b - a)/n$. The endpoints of the partitions segments are $x_k = a + k \Delta x$ with $k = 0, 1, 2, \ldots, n$, such that $x_0 = a$ and $x_n = b$. Let $x_k^*$ be a sample point in the interval $[x_{k-1}, x_k]$.
Definition 5.2. The area $A$ of the region that lies under the graph of a continuous function $f(x) \geq 0$ on an interval $[a, b]$ is
\[
A = \lim_{n \to \infty} A_n^* = \lim_{n \to \infty} \left[ f(x_1^*) \Delta x + f(x_2^*) \Delta x + \cdots + f(x_n^*) \Delta x \right]
\]
for any choice of sample points $x_k^*$.

Let us assess this definition. Any continuous function attains its maximum and minimum values on a closed interval. Let $M_k$ and $m_k$ be, respectively, the maximum and minimum values of $f(x)$ on the interval $[x_{k-1}, x_k]$. If $S_k$ is the area under the graph $y = f(x)$ on the interval $[x_{k-1}, x_k]$, then $S_k^L = m_k \Delta x \leq S_k \leq S_k^U = M_k \Delta x$. The area $S_k^* = f(x_k^*) \Delta x$ of the rectangle with base $\Delta x$ and height $f(x_k^*)$ is a continuous function of $x_k^*$ on the interval $[x_{k-1}, x_k]$. Therefore, $S_k^*$ must take all the values between its minimum and maximum values, $S_k^L$ and $S_k^U$. In particular, $S_k^* = S_k$ for some $x_k^* \in [x_{k-1}, x_k]$. Thus, for any fixed $n$, there is a choice of sample points such that $A_n^* = A$.

Continuing the analogy with the example of $f(x) = x^2$, let us show that the limit (5.2) is independent of the choice of sample points, provided the lower sums $A_n^L = S_1^L + \cdots + S_n^L$ and the upper sums $A_n^U = S_1^U + \cdots + S_n^U$ converge to the same number as $n \to \infty$. Indeed, for any choice of sample points $S_k^L \leq S_k \leq S_k^U$ and, hence, by taking the sum over the partition in the latter inequality, one infers that $A_n^L \leq A_n^* \leq A_n^U$. Therefore both the numbers $A_n^*$ and $A$ lie between $A_n^L$ and $A_n^U$:
\[
\begin{cases}
A_n^L \leq A_n^* \leq A_n^U \\
A_n^L \leq A \leq A_n^U
\end{cases} \implies |A_n^* - A| \leq A_n^U - A_n^L
\]
Thus, if $A_n^U - A_n^L \to 0$ as $n \to \infty$, then $A_n^* \to A$ for any choice of partition. The following theorem holds.

Theorem 5.1. Let $f$ be a continuous function on $[a, b]$. Suppose that for any partition $x_0 = a < x_1 < x_2 < \cdots < x_n = b$, the length of the largest partition interval, $\Delta_n = \max_k \Delta x_k$, $\Delta x_k = x_k - x_{k-1}$, $k = 1, 2, \ldots, n$, decreases as the number $n$ of partition intervals increases, $\Delta_n > \Delta_{n+1}$. Then the upper and lower sums converge to the same limit as $n \to \infty$.

This theorem justifies the definition (5.2). Note also that the partition is not generally required to be equispaced. The above theorem only requires that the length $\Delta_n$ of the largest partition interval decreases with increasing the number of partition intervals ($\Delta_n = (b-a)/n = \Delta x$ for an equispaced partition).
31.4. Approximating the Area Under a Graph. In practice, Equation (5.2) can be used to find the area under the graph that is correct to any desired number of decimal places. Take a partition of the interval \([a, b]\), e.g., fix some \(n\) so that \(\Delta x = (b - a)/n\). Choose sample points \(x_k - 1 \leq x_k^* \leq x_k\). Convenient choices might be the left points \(x_k^* = x_{k-1}\), the right points \(x_k^* = x_k\), or the midpoints \(x_k^* = (x_{k-1} + x_k)/2\). Calculate the sum \(A_n^*\), keeping the desired number of decimal places. Refine the partition by, for example, doubling the number of segments, and calculate \(A_{2n}^*\). If \(A_n^*\) and \(A_{2n}^*\) coincide in the desired number of decimal places, then \(A = A_{2n}^*\) is correct to that number of decimal places. If not, refine the partition further and compute \(A_{4n}^*\) and compare it with \(A_{2n}^*\) and so on, until the needed accuracy is reached. For any \(n\), the absolute error of the approximation may be estimated by the inequality on the right in (5.3).

31.5. Sigma Notation for Sums. To avoid writing lengthy expressions for sums of an arbitrary number of terms, it is convenient to adopt the following notation:

\[
A_n^* = S_1^* + S_2^* + \cdots + S_n^* = \sum_{k=1}^{n} S_k^* ,
\]

where the index \(k\) is called the summation index. The symbol \(\sum\) means adding all \(S_k^*\), starting with \(k = 1\) up to \(k = n\). For example, the geometric sum formula can now be written as

\[
1 + q + q^2 + \cdots + q^n = \sum_{k=0}^{n} q^k = \frac{q^{n+1} - 1}{q - 1} .
\]

31.6. The Distance Problem. If an object moves with a constant velocity \(v\) during a time interval \(a \leq t \leq b\), then the distance traveled by the object is \(D = v(b - a)\). How does one calculate the distance if the speed is a nonconstant continuous function of time \(v = v(t) \geq 0\)?

Let \(D(t)\) be the distance as a function of time \(a \leq t \leq b\). It satisfies the condition \(D(a) = 0\). Since \(v(t) \geq 0\), the object travels in the same direction all the time, and \(D(t)\) increases because \(D'(t) = v(t) \geq 0\). Thus, \(D = D(b)\). To calculate \(D(b)\), consider a partition of \([a, b]\) by interval \([t_{k-1}, t_k]\) where \(t_k = a + \Delta t k\), \(\Delta t = (b - a)/n\), \(k = 0, 1, \ldots, n\). The distance \(\Delta D_k = D(t_k) - D(t_{k-1})\) traveled by the object in the time interval \([t_{k-1}, t_k]\) can be found by the mean value theorem: \(D(t_k) - D(t_{k-1}) = v(t_k^*) \Delta t\) for some \(t_k^*\) in \([t_{k-1}, t_k]\). Recall that \(v(t_k^*)\) is the average velocity over the time interval \([t_{k-1}, t_k]\). The total distance is \(D = \Delta D_1 + \cdots + \Delta D_n\). On the other hand, points
$t_k^*$ represent a particular choice of sample points in the definition (5.2) applied to a continuous function $v(t)$. Therefore, $D$ is the area under the graph of $v(t)$ and, hence, can be calculated with any choice of sample points $t_k^*$, not necessarily with those at which $v$ coincides with the average velocity in each partition interval:

$$D = \lim_{n \to \infty} \sum_{k=1}^{n} v(t_k^*) \Delta t,$$

Furthermore, by the condition $D'(t) = v(t)$ the function $D(t)$ is the antiderivative of $v(t)$ satisfying the initial condition $D(a) = 0$. If $F(t)$ is a particular antiderivative of $v(t)$, then $D(t)$ and $F(t)$ can differ only by a constant, $D(t) = F(t) + C$ for $a < t < b$. The constant $C$ is fixed by the condition $D(a) = 0$ and, hence, $C = -F(a)$. The distance traveled is $D = D(b) = F(b) - F(a)$ This establishes the following relation between the area under the graph of $v(t)$ and an antiderivative $F(t)$ of $v(t)$:

(5.5) $$D = \lim_{n \to \infty} \sum_{k=1}^{n} v(t_k^*) \Delta t = F(b) - F(a)$$

**Example 5.1.** A moving object slows down so that its velocity is $v(t) = e^{-2t}$. What is the distance traveled by the object during the time interval $0 \leq t \leq 1$?

**Solution:** Let $\Delta t = 1/n$ so that $t_k = k/n$, $k = 0, 1, \ldots, n$. Take $t_k^* = (k-1)/n$, $k = 1, 2, \ldots, n$ (the left points of partition intervals). Then $v(t_k^*) \Delta t = q^{k-1}/n$, where $q = e^{-2/n}$. The distance traveled is

$$D = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} q^k = \lim_{n \to \infty} \frac{1}{n} \frac{q^n - 1}{q - 1} = \frac{1 - e^{-2}}{\lim_{n \to \infty} n(1 - e^{-2/n})},$$

where the sum formula (5.4) has been used. To compute the limit in the denominator, let $x = 1/n$, that is, $x \to 0^+$. The limit becomes the indeterminate form $(1 - e^{-2x})/x$ of type $0/0$, which can be resolved by l’Hospital’s rule:

$$\lim_{n \to \infty} n(1 - e^{-2/n}) = \lim_{x \to 0^+} \frac{1 - e^{-2x}}{x} = \lim_{x \to 0^+} \frac{(1 - e^{-2x})'}{(x)'} = \lim_{x \to 0^+} \frac{2e^{-2x}}{1} = 2$$
Thus, the distance traveled is \( D = (1 - e^{-2})/2 \).

**Alternative solution:** Using the table of antiderivatives

\[
F'(t) = v(t) = e^{-2t} \implies F(t) = -\frac{1}{2}e^{-2t}.
\]

By Equation (5.5), \( D = F(1) - F(0) = (1 - e^{-2})/2 \). Note that any particular antiderivative can be used.

When compared to the previous solution, this one looks like cheating! More to the point, take \( v(t) = t^2 \) (the example discussed at the beginning of this section). Its particular antiderivative is \( F(t) = t^3/3 \). So the distance traveled, or the area under the graph of \( t^2 \), is \( F(1) - F(0) = 1/3 \). It turns out that the relation (5.5) between an antiderivative of a function and the area under the graph of the function is not specific for the distance problem. Its generalization will be established with the help of the concept of the **definite integral**.

### 31.7. Exercises.

1. Find explicit formulas for the upper and lower sums, \( A_n^U \) and \( A_n^L \), for \( f(x) = 2x + 1 \) on \([0, 2]\) using an equispaced partition. Find the limits of \( A_n^U \) and \( A_n^L \) as \( n \to \infty \). What is the geometrical significance of this limit?

2. Find explicit formulas for the upper and lower sums, \( A_n^U \) and \( A_n^L \), for \( f(x) = x^3 \) on \([0, 1]\) using an equispaced partition. Show that \( A_n^{U,L} \to 1/4 \) as \( n \to \infty \). What is the geometrical significance of this limit? *Hint:*

\[
\sum_{k=1}^{n} k^3 = \frac{1}{4} n^2(n - 1)^2.
\]

3. Find the area under the graph of \( f(x) = e^{-x^2} \) on \([-1, 1]\) correct up to five decimal places.

4. Find the area under the graph \( f(x) = \sqrt{1 - x^2} \), where \(-1 \leq x \leq 1\), correct up to three decimal places. Use the geometrical interpretation of this area to find its exact value.

5. Find the area under the graph of each of the following functions on the given interval using the relation (5.2):
   - (i) \( f(x) = 3 - 3x \), \( 0 \leq x \leq 1 \)
   - (ii) \( f(x) = 1 + x + x^2 \), \( 0 \leq x \leq 2 \)
   - (iii) \( f(x) = e^{3x} \), \(-1 \leq x \leq 1 \)
(6) Use the relation
\[ \sum_{k=0}^{n-1} \sin(2kx) = \frac{\sin(nx) \sin((n-1)x)}{\sin x} \]
to find the following:
(i) The upper and lower sums for \( f(x) = \sin x \) on the interval \([0, \pi]\). Calculate \( A_n^U - A_n^L \) and investigate its behavior as \( n \) increases. What is the significance of this number for a fixed \( n \)?

(ii) The area under the graph of \( f(x) = \sin x \), \( 0 \leq x \leq \pi \), using (5.2).

(iii) The area under the graph of \( f(x) = \cos x \), \( 0 \leq x \leq \pi/2 \), using (5.2).

(7) An object travels with velocity \( v(t) = \cos^2 t \). Find the distance passed by the object over the time interval \( 0 \leq t \leq 2\pi \).

(8) Use the table of antiderivatives to find the area under the graph of each of the following functions. Sketch the graph of the function on a given interval and explain why this method can be used to find the area.
(i) \( f(x) = 1/(x^2 + 1), -1 \leq x \leq 1 \)
(ii) \( f(x) = \sin(ax), 0 \leq x \leq \pi/a \)
(iii) \( f(x) = \sqrt{x}, 0 \leq x \leq 4 \)
(iv) \( f(x) = 1/x, 1 \leq x \leq 2 \)
(v) \( f(x) = \tan x, \pi/4 \leq x \leq \pi/3 \)
(vi) \( f(x) = 1/\sqrt{x^2 + 1}, -1 \leq x \leq 1 \)
(vii) \( f(x) = 2^x, -2 \leq x \leq 2 \)
(viii) \( f(x) = x/(x^2 + 1), 0 \leq x \leq 1 \)
(ix) \( f(x) = x^{4/3}, -27 \leq x \leq 27 \)
(x) \( f(x) = 1/(1 - x^2), -1/2 \leq x \leq 1/3 \)

(9) A car starts accelerating at \( t = 0 \) so that its speed is \( v = at \), \( a > 0 \). At a time \( t = b \), it begins to slow down so that its speed becomes \( v(t) = ab + c(b - t), c > 0 \), until its speed vanishes and the car stops. Find the distance traveled by the car. \textit{Hint:} Sketch the graph of \( v(t) \) in the time interval in which the car was moving.

(10) An object travels with speed \( v(t) = \sqrt{a^2 - t^2} \), starting at \( t = -a \) and stopping at \( t = a \). Sketch the graph of \( v(t) \) and find the distance traveled by the object.

(11) Let \( f(x) = (x^5 - 1)/(x - 1) \) if \( 0 \leq x < 1 \) and \( f(1) = 5 \). Show that \( f(x) \) is continuous on \([0, 1]\). Use antiderivatives to find the area under the graph of \( f \). \textit{Hint:} See Equation (5.4).
(12) Find the area of a planar region bounded by the curves $y = 2 - x^2$ and $y = 1$. Sketch the region.
(13) Find the area of a planar region bounded by the curves $y = x^2$ and $y = x$. Sketch the region.

32. The Definite Integral

A generalization of the concept of the area under a graph leads to one of the most fundamental concepts in calculus, the definite integral.

32.1. Supremum and Infimum. The area under a graph is also well defined if the function has some number of bounded jump discontinuities. The difference with the case of a continuous function $f$ is that now $f$ may or may not attain its maximum or minimum values on each partition interval. What should be changed in the definition of the area to accommodate possible jump discontinuities of the graph? Suppose a function $f$ is bounded on an interval $[a, b]$; that is, there are numbers $m$ and $M$ such that $m \leq f(x) \leq M$ for all $x \in [a, b]$. If $m$ is a lower bound, then any number $m_1 < m$ is also a lower bound, but a number $m_2 > m$ may or may not be a lower bound. So one can find the greatest lower bound that is unique for $f$ on $[a, b]$. Similarly, one can find the least upper bound of $f$ on $[a, b]$. These bounds have special names.

Definition 5.3. (Infimum and Supremum). The number $m$ is called the infimum of a bounded function $f$ on an interval $I = [a, b]$ if $m$ is a lower bound of $f$ but $m + \varepsilon$ is not a lower bound for any $\varepsilon > 0$. This number is denoted as $m = \inf I f$. The number $M$ is called the supremum of $f$ on $[a, b]$ if $M$ is an upper bound of $f$ but $M - \varepsilon$ is not an upper bound for any $\varepsilon > 0$. This number is denoted as $M = \sup I f$.

Remark. The completeness axiom for a set of real numbers says that if $S$ is a nonempty set of real numbers that has an upper bound $M$, then $S$ has the least upper bound $\inf S$. If $S$ has a lower bound $m$, then it also has the greatest lower bound $\inf S$. The completeness axiom is an expression of the fact that there is no gap or hole in the real number line. The numbers $\sup S$ and $\inf S$ are unique. Indeed, assume that $M_1 \neq M_2$ are the least upper bounds of $S$. Since they are not equal one of them should be less than the other, e.g., $M_1 < M_2$. Since $M_2$ is the least upper bound, the number $M_2 - \varepsilon$ is not an upper bound for any $\varepsilon$. Take $\varepsilon = M_2 - M_1 > 0$. Then $M_2 - \varepsilon = M_1$ but $M_1$ is also an upper bound, hence, a contradiction. The uniqueness of $\inf S$ is established by similar lines of reasoning.
Naturally, if the function is continuous, then \( \sup f \) is nothing but the maximum value of \( f \) and \( \inf f \) is its minimum value. However, if a function has jump discontinuities, then \( \sup f \) and \( \inf f \) always exist, while the maximum and minimum values may not exist. This is illustrated in Figure 5.2.

32.2. Definition of the Definite Integral. Let \( f \) be a bounded function on an interval \([a, b]\). Consider a partition of \([a, b]\) by \( n \) intervals \( I_k = [x_{k-1}, x_k] \), \( k = 1, 2, \ldots, n \), where \( a = x_0 < x_1 < \cdots < x_n = b \). Let \( M_k = \sup_{I_k} f(x) \) (the supremum of \( f(x) \) on the interval \( I_k \)) and \( m_k = \inf_{I_k} f(x) \) (the infimum of \( f(x) \) in the interval \( I_k \)). The length of \( I_k \) is \( \Delta x_k = x_k - x_{k-1} \). The lower \( A_n^L \) and upper \( A_n^U \) sums for \( f \) are defined by

\[
A_n^L = \sum_{k=1}^{n} m_k \Delta x_k, \quad A_n^U = \sum_{k=1}^{n} M_k \Delta x_k
\]

for every partition of \([a, b]\). Put \( \Delta_n = \max_k \Delta x_k \) which is the length of the largest partition interval for a fixed \( n \). The sequences of lower and upper sum are defined so that \( \Delta_n > \Delta_{n+1} \); that is, the length of the largest partition interval decreases with increasing the number of partition intervals.

Definition 5.4 (The Definite Integral). A bounded function \( f \) is said to be integrable on an interval \([a, b]\) if the sequences of its lower and upper sums converge to the same number. This number is called the definite integral of \( f \) from \( a \) to \( b \) and is denoted by

\[
\int_a^b f(x) \, dx = \lim_{n \to \infty} A_n^L = \lim_{n \to \infty} A_n^U;
\]

the numbers \( a \) and \( b \) are called the lower and upper integration limits, respectively, and the function \( f \) is called the integrand.

Apparently, for a continuous and nonnegative \( f \) on \([a, b]\), the definite integral coincides with the area under the graph of \( f \). Similarly to the area under the graph of a continuous nonnegative function, an integrable function has the property

\[
A_n^L \leq \int_a^b f(x) \, dx \leq A_n^U
\]

for any \( n \) (see Exercise 32.9.4).

32.3. Riemann Sums. There is an analog of Equation (5.2) for the definite integral.
Figure 5.2. Relations between the supremum and infimum of $f$ and the maximum and minimum values of $f$.

**Upper left panel**: The values of the function approach 1 as $x$ approaches $c$ from the left, but $f(c) = 1/2 < 1$. The maximum value of $f$ does not exist, but the least upper bound does exist, $\sup f = 1$.

**Lower left panel**: The values of $f$ approach $1/2$ as $x$ approaches $c$ from the right, but $f(c) = 1$. The function has no minimum value, but the greatest lower bound is $\inf f = 1/2$.

**Upper right panel**: The values of $f$ approach 1 as $x$ approaches $c$ from the left and $f(c) = 1$. In this case, the maximal value $f(c) = 1$ coincides with $\sup f = 1$.

**Lower right panel**: The values of $f$ approach $1/2$ as $x$ approaches $c$ from the right and $f(c) = 1/2$. The minimum value $f(c) = 1/2$ coincides with the greatest lower bound $\inf f = 1/2$. 
DEFINITION 5.5. Let $I_k$ be partition intervals of $[a, b]$, $\Delta x_k$ be the length of $I_k$, and $x_k^* \in I_k$. The sum

$$R_n(f) = \sum_{k=1}^{n} f(x_k^*) \Delta x_k$$

is called a Riemann sum of a function $f$ on $[a, b]$.

The sum $R_n(f)$ is named after the German mathematician Bernhard Riemann (1826–1866). Evidently, the value of the Riemann sum generally depends on the choice of partition intervals and sample points $x_k^*$. However, for integrable functions Riemann sums have a remarkable property. Let a sequence of Riemann sums $R_n$ be defined similarly to the sequences of the upper and lower sums, i.e., the largest partition segment decreases with increasing $n$.

THEOREM 5.2. If $f$ is integrable on $[a, b]$, then, for any number $\varepsilon > 0$, there exists an integer $N$ such that

$$\left| \int_{a}^{b} f(x) \, dx - R_n(f) \right| < \varepsilon$$

for every integer $n > N$ and for every choice of $x_k^*$ in $I_k$.

A proof of this theorem is given as an exercise (see Exercise 32.9.5; see also Exercises 32.9.12 and 32.9.13). The theorem asserts that a Riemann sum for a sufficiently large $n$ can approximate the definite integral with any desired accuracy; that is, for any (small) designated absolute error $\varepsilon$, $R_n(f)$ differs from $\int_{a}^{b} f(x) \, dx$ no more than $\varepsilon$ for a sufficiently large $n$. In other words,

$$(5.6) \quad \lim_{n \to \infty} R_n(f) = \int_{a}^{b} f(x) \, dx,$$

for any choice of sample points $x_k^*$. Equation (5.6) is the analog of Equation (5.2). It can be understood from the inequality $A_n^L \leq R_n(f) \leq A_n^U$, which follows from $m_k \leq f(x_k^*) \leq M_k$ for any $x_k^*$ (see Figure 5.3).

For an integrable function, $A_n^L$ and $A_n^U$ converge to the same number, which is the value of the definite integral, and, by the squeeze principle, so should $R_n(f)$ independently of the choice of sample points.

32.4. Continuity and Integrability. The relation (5.6) can be used to calculate the definite integral, provided the function $f$ is integrable. The question of integrability requires investigating the convergence of the sequences of the upper and lower sums, which might be a tedious task even for such simple functions as, for example, $f(x) = x^2$, as
discussed in the previous section. The following theorem is helpful when studying the question of integrability.

**Theorem 5.3.** If $f$ is continuous on $[a, b]$, or if $f$ has only a finite number of bounded jump discontinuities, then $f$ is integrable on $[a, b]$; that is, the definite integral $\int_a^b f(x) \, dx$ exists.

**An Example of a Nonintegrable Function.** A bounded function $f$ with infinitely many jump discontinuities may or may not be integrable. So, in general, the area under the graph of such a function cannot be unambiguously defined. As an example, consider a bounded nonnegative function $f$ on $[0, 1]$ such that

$$f(x) = \begin{cases} 
1 & \text{if } x \text{ is a rational number} \\
0 & \text{if } x \text{ is an irrational number}
\end{cases}$$

The function is not continuous anywhere in $[0, 1]$ and has infinitely many jump discontinuities. For example, $f(1/2) = 1$, but when $x$ approaches $1/2$, the value $f(x)$ keeps jumping from 0 to 1 and back, no matter how close $x$ is to $1/2$ because, for any $\delta > 0$, the interval...
\(\frac{1}{2} - \delta, \frac{1}{2} + \delta\) always contains both rational and irrational numbers. This function is not integrable. Indeed, take a partition \(x_k = k/n, k = 0, 1, \ldots, n\). Any partition interval \([(k - 1)/n, k/n]\) contains both rational and irrational numbers. Therefore, \(m_k = 0\) and \(M_k = 1\). Hence, the lower sum vanishes for any partition, \(A^L_n = 0\), whereas the upper sum is \(A^U_n = \sum_{k=1}^{n} \Delta x = 1\), that is, \(\lim_{n \to \infty} A^L_n = 0\) while \(\lim_{n \to \infty} A^U_n = 1\). The function is not integrable. The integral does not exist. Note that the Riemann sum can still be defined, but its limit would depend on the choice of sample points (e.g., take \(x^*_k\) to be rational numbers or take \(x^*_k\) to be irrational numbers; both options are possible since any partition interval always contains rational and irrational numbers). In fact, with a suitable choice of sample points, the Riemann sums can converge to any value between 0 and 1 (e.g., in all partition segments to left of \(0 < p < 1\), take \(x^*_k\) to be rational, while in all partition segments to the right of \(p\), take \(x^*_k\) to be irrational).

32.5. Properties of the Definite Integral. Suppose \(f(x) = c\), where \(c\) is a constant. In this case, for any partition interval \(I_k\), \(M_k = m_k = c\) and \(A^L_n = A^U_n = c\sum_{k=1}^{n} \Delta x = cn \Delta x = c(b - a)\). In other words, a constant function is integrable and its integral is \(c(b - a)\):

\[
\int_{a}^{b} c\, dx = c(b - a). \tag{5.7}
\]

For any two integrable functions \(f(x)\) and \(g(x)\) and constants \(c_1\) and \(c_2\), it follows from the convergence of the Riemann sums (5.6) for \(f\) and \(g\) that

\[
\int_{a}^{b} \left[ c_1 f(x) + c_2 g(x) \right] \, dx = \lim_{n \to \infty} \sum_{k=1}^{n} \left[ c_1 f(x^*_k) + c_2 g(x^*_k) \right] \Delta x_k
\]

\[
= c_1 \lim_{n \to \infty} \sum_{k=1}^{n} f(x^*_k) \Delta x_k + c_2 \lim_{n \to \infty} \sum_{k=1}^{n} g(x^*_k) \Delta x_k
\]

\[
= c_1 \int_{a}^{b} f(x) \, dx + c_2 \int_{a}^{b} g(x) \, dx. \tag{5.8}
\]

So the integration is a linear operation. In particular, the integral of the sum of two functions is the sum of their integrals. The integral of a function multiplied by a constant is the product of the constant and the integral of the function. If the integration limits are reversed, then all \(\Delta x_k\) change their signs as \(x_k\) becomes less than \(x_{k-1}\). Therefore,

\[
\int_{a}^{b} f(x) \, dx = - \int_{b}^{a} f(x) \, dx. \tag{5.9}
\]
and, in particular,

\[(5.10) \int_a^a f(x) \, dx = 0.\]

It can be proved that

\[(5.11) \int_a^b f(x) \, dx = \int_a^c f(x) \, dx + \int_c^b f(x) \, dx\]

for \(f\) integrable on \([a, b]\) and any \(a \leq c \leq b\). The proof is rather technical and is omitted. If \(f\) is continuous and positive on \([a, b]\), then the property \((5.11)\) is trivial: The area under the graph of \(f\) on \([a, b]\) is the sum of the areas under the graph of \(f\) on \([a, c]\) and \([c, b]\).

### 32.6. Geometrical Significance of the Definite Integral

As already noted, the definite integral of \(f\) from \(a\) to \(b\) coincides with the area under the graph of \(f\) for a continuous and positive \(f\). Suppose \(f\) is continuous and negative on \([a, b]\). Consider the function \(g(x) = -f(x)\). The integral of \(g\) is the area \(A\) under the graph of \(g\) and, hence, \(A\) also coincides with the area above the graph of \(f\) and below the \(x\) axis. By the linearity of the integral, \(\int_a^b f(x) \, dx = -\int_a^b g(x) \, dx = -A\). So, for a negative \(f\), the integral of \(f\) coincides with the negative area of the region bounded below by the graph of \(f\) and above by the \(x\) axis. Now let \(f\) be continuous on \([a, b]\). Let it be positive on \([a, c]\) and negative on \([c, b]\), that is, \(f(c) = 0\). Then it follows from the property \((5.11)\) that

\[\int_a^b f(x) \, dx = \int_a^c f(x) \, dx + \int_c^b f(x) \, dx = A_1 - A_2,\]

where \(A_1\) is the area under the graph of \(f\) on \([a, c]\) and \(A_2\) is the area above the graph of \(f\) on \([c, b]\). This property is illustrated in Figure 5.4.

### 32.7. Comparison Properties of the Integral

The following additional properties of the definite integral can be established:

\[(5.12) \int_a^b f(x) \, dx \geq 0, \quad \text{if } f(x) \geq 0 \text{ in } [a, b],\]

\[(5.13) \int_a^b f(x) \, dx \geq \int_a^b g(x) \, dx, \quad \text{if } f(x) \geq g(x) \text{ in } [a, b],\]

\[(5.14) m(b - a) \leq \int_a^b f(x) \, dx \leq M(b - a), \quad \text{if } m \leq f(x) \leq M \text{ in } [a, b].\]

The property \((5.12)\) follows directly from the definition. Since \(0 \leq m_k \leq M_k\) for any partition if \(f(x) \geq 0\), the upper and lower sums
are nonnegative and so must be the integral. If $f$ is continuous, the property (5.12) states the obvious that the area under the graph of $f$ is nonnegative. The property (5.13) follows from (5.12) for the function $f(x) - g(x) \geq 0$ and the linearity of the integral (5.8). The property (5.14) is also a consequence of the definition. Indeed, for any partition, $m \leq m_k \leq M_k \leq M$. Hence, $m(b-a) \leq A_n^L \leq A_n^U \leq M(b-a)$ for any $n$. In the limit $n \to \infty$, this inequality turns into (5.14).

**Theorem 5.4 (Integrability of the absolute value).** If a bounded function $f(x)$ is integrable over an interval $[a, b]$, then its absolute value $|f(x)|$ is also integrable over $[a, b]$ and

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx$$

A proof of this theorem under a simplifying hypothesis that $f$ is continuous is given as an exercise (Exercise 32.9.11). The converse of Theorem 5.4 is not true. The integrability of the absolute value $|f(x)|$
Figure 5.5. Geometrical interpretation of the property (5.14). The graph of a function $f$ lies between two horizontal lines $y = m$ and $y = M$ because $m \leq f(x) \leq M$ for all $x \in [a, b]$. So the area $A$ under the graph of $f$ lies between the areas of rectangles with the base $b - a$ and heights $m$ and $M$, i.e., $m(b - a) \leq A \leq M(b - a)$.

does not imply the integrability of $f(x)$. Consider the function

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is a rational number} \\ -1 & \text{if } x \text{ is a irrational number} \end{cases}$$

The absolute value $|f(x)| = 1$ is continuous and, hence, is integrable on any bounded interval. However, $f(x)$ is not integrable. The proof of this assertion is left to the reader as an exercise.

32.8. Evaluation of the Integral by the Riemann Sum. If the integral exists ($f$ is integrable), then it can be evaluated as the limit of the Riemann sum (5.6). The limit is independent of the choice of sample points. The following choices are often used in practice:

$$x_k^* = x_{k-1} \quad \text{(the left-point rule),}$$
$$x_k^* = x_k \quad \text{(the right-point rule),}$$
$$x_k^* = (x_{k-1} + x_k)/2 \quad \text{(the midpoint rule),}$$
in combination with the basic properties of the integral. The evaluation of the Riemann sum is rather technical. Formulas like (5.1), (5.4), and (5.15)

$$\sum_{k=1}^{n} k = \frac{n(n - 1)}{2}, \quad \sum_{k=1}^{n} k^3 = \left[ \frac{n(n - 1)}{2} \right]^2$$
can be helpful. However, the Riemann sum is mostly used to calculate the integral *approximately* with some designated accuracy by means of computer simulations, similarly to approximate calculations of the area discussed in the previous section.

**Example 5.2.** Find the definite integral of \( f(x) = e^{-2x} - 2x^2 + 4x^3 \) from 0 to 1.

**Solution: 1.** The function is continuous on \([0, 1]\) and hence integrable; that is, Equation (5.6) applies for any choice of \( x_k^* \). The left-point rule will be used.

2. By the linearity of the integral,

\[
\int_0^1 f(x) \, dx = \int_0^1 e^{-2x} \, dx - 2 \int_0^1 x^2 \, dx + 4 \int_0^1 x^3 \, dx.
\]

The first integral is \((1 - e^{-2})/2\) by Example 5.1 (where the area under the graph of \( e^{-2x} \) in \([0, 1]\) was calculated). The area under the graph \( x^2 \) in \([0, 1]\) can be found at the beginning of the previous section and is equal to 1/3. The area under the graph of \( x^3 \) can be found with the help of the second relation in (5.15). Let \( \Delta x = 1/n \) and \( x_k = (k-1)/n \) (the left-point rule), then the Riemann sum (5.6) becomes

\[
\int_0^1 x^3 \, dx = \lim_{n \to \infty} \frac{1}{n^4} \sum_{k=1}^{n} k^3 = \lim_{n \to \infty} \frac{1}{n^4} \frac{n^2(n-1)^2}{4} = \frac{1}{4}.
\]

3. Thus,

\[
\int_0^1 f(x) \, dx = \frac{1 - e^{-2}}{2} - \frac{2}{3} + \frac{1}{4} = \frac{1 - 6e^{-2}}{12}. \quad \square
\]

**32.9. Exercises.**

1. Let \( f(x) = \sin(1/x) \) if \( x \neq 0 \) and \( f(0) = f_0 \). Given any number \( \delta > 0 \), find the supremum and infimum of \( f \) on \([-\delta, \delta]\).

2. Find the upper and lower sums for the function \( f(x) = 1 \) if \( x \geq 0 \) and \( f(x) = -2 \) if \( x < 0 \) on the interval \([-1, 1]\). Use them to show that \( f \) is integrable. Find the value of the integral as the limit of the lower and upper sums.

3. Find the upper and lower sums for the function \( f(x) = 1 \) if \( x \neq 0 \) and \( f(x) = f_0 \neq 1 \) if \( x = 0 \) on an interval \([a, b] \). Use them to show that \( f \) is integrable on any \([a, b] \). Find the value of the integral as the limit of the lower and upper sums.

4. Suppose that the length of the largest partition interval, \( \Delta_n = \max_k \Delta x_k \), decreases as the number \( n \) of partition intervals
increases (i.e., \( \Delta_{n+1} < \Delta_n \)). Show that \( A_n^L \leq A_{n+1}^L \) and \( A_n^U \geq A_{n+1}^U \). Deduce from this property that

\[
A_n^L \leq \int_a^b f \, dx \leq A_n^U
\]

for any \( n \) and for any integrable \( f \) on \([a, b]\).

(5) Use the inequality from the previous exercise to prove Theorem 5.2.

(6) Let \( a < c < b \). Put \( I_1 = [a, c] \) and \( I_2 = [c, b] \). Let \( f \) be integrable on \([a, b]\). If \( M_1 = \sup_{I_1} f, \ M_2 = \sup_{I_2} f, \ m_1 = \inf_{I_1} f, \) and \( m_2 = \inf_{I_2} f \), prove that

\[
m_1(c - a) + m_2(b - c) \leq \int_a^b f \, dx \leq M_1(c - a) + M_2(b - c).
\]

(7) Use Equation (5.14) to estimate the definite integral of each of the following functions from above and below:

(i) \( f(x) = x^5, \ -1 \leq x \leq 2 \)
(ii) \( f(x) = |\sin x| - \sin x, \ 0 \leq x \leq 2\pi \)
(iii) \( f(x) = \cos x - 1/2 - \cos x, \ -\pi \leq x \leq \pi \)

(8) Use the result of exercise 6 to improve the upper and lower estimates in exercise 7 by a suitable choice of \( c \).

(9) Use the geometrical properties of the definite integral to find the exact values of

(i) \( \int_0^1 f(x) \, dx, \ f(x) = \begin{cases} 0, & x = 1/3 \\ 2, & x \neq 1/3 \end{cases} \)
(ii) \( \int_0^3 f(x) \, dx, \ f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 2, & 1 < x < 2 \\ -3, & 2 \leq x \leq 3 \end{cases} \)
(iii) \( \int_0^a f(x) \, dx, \ a > 0, \ f(x) = \begin{cases} x, & x \leq a/2 \\ -2a + x, & x > a/2 \end{cases} \)
(iv) \( \int_{-a}^a \sqrt{a^2 - x^2} \, dx \)
(v) \( \int_{-1/\sqrt{2}}^{1/\sqrt{2}} \left( \sqrt{1-x^2} - |x| \right) \, dx \)
(vi) \( \int_0^a \left[ \sqrt{a^2 - x^2} - \frac{1}{2} \sqrt{a^2 - (2x-a)^2} \right] \, dx \)

(10) Use Riemann sums for equispaced partitions to evaluate each of the following definite integrals correct up to three decimal places:
(i) \( \int_{-1}^{2} x^{1/3} \, dx \)
(ii) \( \int_{0}^{2} \sin(x^2) \, dx \)
(iii) \( \int_{-1}^{1} \exp(\sin x) \, dx \)
(iv) \( \int_{1}^{3} x \ln x \, dx \)

(11) Let \( f \) be continuous on \([a, b]\). Then \( g(x) = |f(x)| \) is integrable on \([a, b]\). Why? Show that 
\[
\left| \int_{a}^{b} f \, dx \right| \leq \int_{a}^{b} |f(x)| \, dx.
\]

*Hint:* Compare Riemann sums for \( f(x) \) and \(|f(x)|\).

(12) Let \( f \) have a bounded derivative on \([a, b]\), that is, \(|f'(x)| \leq M_1\) for all \( x \in [a, b] \). Consider an equispaced partition of \([a, b]\) with \( \Delta x = (b - a)/n \). For every partition interval \( I_k = [x_{k-1}, x_k] \), \( x_k = a + kx \Delta x \), \( k = 0, 1, \ldots, n \), show that there is \( \bar{x}_k \in I_k \) such that \( A_k = f(\bar{x}_k)\Delta x \) is the area under the graph of \( f \) on \( I_k \). Let \( x_k^* \in I_k \) be sample points in the Riemann sum for \( f \). Use the mean value theorem to prove that 
\[
|f(x_k^*) \Delta x - A_k| \leq M_1 \Delta x^2
\]
for any \( k \). Deduce from this inequality that 
\[
\left| \int_{a}^{b} f \, dx - R_n(f) \right| \leq \frac{M_1(b - a)^2}{n}.
\]

(13) (The Trapezoidal Rule). Let \( f \) have a bounded second derivative on \([a, b]\), that is, \(|f''(x)| \leq M_2\) for all \( x \in [a, b] \). By the mean value theorem, there is \( c \) such that \( f(b) - f(a) = f'(c)(b - a) \). Define the function \( g(x) = f(a) + f'(c)(x - a) \). The graph \( y = g(x) \) is the secant line through the points \((a, f(a))\) and \((b, f(b))\). Then \( T = \int_{a}^{b} g(x) \, dx = (f(a) + f(b))(b - a)/2 \) is the area of the trapezoid bounded by the line \( y = g(x) \) on \([a, b]\). Use the mean value theorem for the derivative \( f' \) to prove that 
\[
|f(x) - g(x)| \leq M_2(b - a)^2.
\]

Use this inequality to show that 
\[
\left| \int_{a}^{b} f \, dx - \int_{a}^{b} g \, dx \right| \leq M_2(b - a)^3.
\]

The trapezoidal rule to calculate \( \int_{a}^{b} f \, dx \) uses the piecewise linear approximation of \( f \) by \( g \) on each partition interval \( I_k = \]}
5. INTEGRATION

\[ [x_{k-1}, x_k] \] of length \( \Delta x_k \):

\[
\int_a^b f \, dx \approx T_n(f) = \sum_{k=1}^{n} \int_{x_{k-1}}^{x_k} g \, dx
\]

\[
= \sum_{k=1}^{n} \frac{1}{2} (f(x_{k-1}) + f(x_k)) \Delta x_k.
\]

Prove that, for an equispaced partition,

\[
\left| \int_a^b f \, dx - T_n(f) \right| \leq \frac{M_2(b-a)^3}{n^2}.
\]

By comparing this result with that in exercise 12, one can see that the error of the trapezoidal rule decreases faster than that in the Riemann sum approximation as the number of partition intervals increases. So it is a better way to approximate the integral. One should keep in mind, however, that the integrand has to have a bounded second derivative for such a superiority.

(14) Evaluate \( \int_0^2 \sin(x^2) \, dx \) correct up to three decimal places using the Riemann sum and trapezoidal approximations. How many partition intervals are required to achieve this accuracy in each of the approximations?

33. The Fundamental Theorem of Calculus

In this section, the relation between the definite integral of a function and its antiderivative will be established. This relation provides a powerful method for calculating the definite integral that avoids the use of Riemann sums.

33.1. Integration and Differentiation. Consider the definite integral of \( f(t) = t \) from 0 to \( x \) for some \( x > 0 \). This integral represents the area under the graph of \( f(t) = t \) in the interval \([0, x]\), which is the area of a right triangle:

\[
A(x) = \int_0^x t \, dt = \frac{x^2}{2}.
\]

The area \( A(x) \) can be viewed as a function of the variable \( x \), which is the length of the triangle catheti. This function has an interesting property:

\[
A'(x) = x = f(x).
\]

In other words, the derivative of the definite integral with respect to its upper limit equals the value of the integrand at the upper limit. Recall that if \( v(t) \geq 0 \) is the speed of a moving object, then the distance
traveled by the object in time $T$ is given by the area under the graph of $v(t)$:

$$D(T) = \int_0^T v(t) \, dt.$$ 

On the other hand, the speed is the rate of change of $D(T)$, and therefore there should be $D'(T) = v(T)$; that is, the derivative of the integral with respect to its upper limit is again the value of the integrand at the upper limit. How general is this property? Does it hold for all integrable functions? The following theorem answers these questions.

**Theorem 5.5.** If $f$ is continuous on $[a, b]$, then the function defined by

$$g(x) = \int_a^x f(t) \, dt, \quad a \leq x \leq b,$$

is continuous on $[a, b]$ and differentiable on $(a, b)$, and $g'(x) = f(x)$.

**Proof.** By the definition of the derivative, one has to prove that

$$(5.16) \quad \lim_{h \to 0} \frac{g(x + h) - g(x)}{h} = f(x)$$

for $a < x < b$. The ratio in the limit can be transformed as follows:

$$\frac{g(x + h) - g(x)}{h} = \frac{1}{h} \left[ \int_a^{x+h} f(t) \, dt - \int_a^x f(t) \, dt \right]$$

$$= \frac{1}{h} \left[ \left( \int_a^x f(t) \, dt + \int_x^{x+h} f(t) \, dt \right) - \int_a^x f(t) \, dt \right]$$

$$= \frac{1}{h} \int_x^{x+h} f(t) \, dt,$$

where the property (5.11) has been used. Note that since $a < x < b$ (i.e., $x \neq a$ and $x \neq b$), for a sufficiently small $h \neq 0$, both $x$ and $x + h$ ($h$ can be positive or negative) always lie in the interval $(a, b)$ so that the interval $[x, x + h]$ is contained in $(a, b)$. By the continuity of $f(t)$ on the interval $[x, x + h]$, the function $f$ attains its absolute maximum and minimum values in $[x, x + h]$. Let $M = f(v)$ and $m = f(u)$ be the absolute maximum and minimum values, respectively, where $v$ and $u$ are in $[x, x + h]$. Suppose that $h > 0$. Then $m \leq f(t) \leq M$ for $x \leq t \leq x + h$ and, by the property (5.14),

$$(5.17) \quad mh = f(u)h \leq \int_x^{x+h} f(t) \, dt \leq f(v)h = Mh.$$
Since $h > 0$, by dividing this inequality by $h$, one can infer that
\begin{equation}
(5.18) \quad f(u) \leq \frac{1}{h} \int_x^{x+h} f(t) \, dt \leq f(v)
\end{equation}
for some $u$ and $v$ in $[x, x + h]$. Inequality (5.18) can be established for $h < 0$ in a similar manner. Indeed, inequality (5.17) holds for the integral $\int_x^{x+h} f(t) \, dt$. After dividing it by $-h > 0$, inequality (5.18) is obtained but with the minus sign at the integral. By the property (5.9), the sign is reversed, yielding (5.18). Thus,
\begin{equation}
(5.19) \quad f(u) \leq \frac{g(x+h) - g(x)}{h} \leq f(v).
\end{equation}
Since $u$ and $v$ lie in the interval $[x, x + h]$,
\[
\lim_{h \to 0} f(u) = f(x), \quad \lim_{h \to 0} f(v) = f(x).
\]
Then the relation (5.16) follows from the squeeze principle:
\[
f(x) = \lim_{h \to 0} f(u) \leq \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \leq \lim_{h \to 0} f(v) = f(x).
\]

This theorem basically states that if a continuous function is first integrated and then differentiated, then it remains unchanged:
\begin{equation}
(5.19) \quad \frac{d}{dx} \int_a^x f(t) \, dt = f(x), \quad a < x < b.
\end{equation}
In other words, $F(x) = \int_a^x f(t) \, dt$ is an antiderivative of $f(x)$ in an open interval $(a, b)$. The continuity of $f$ on $[a, b]$ is essential for this relation to hold. Take, for example, $f(t) = 0$ if $t < 1$ and $f(t) = 1$ if $t \geq 1$. Let $a = 0$. Then $f$ has a jump discontinuity at $t = 1$; it is integrable on any interval, but not continuous at $t = 1$. By the property (5.7), $g(x) = \int_0^x f(t) \, dt = 0$ if $x < 1$. For $x \geq 1$, one has
\[
g(x) = \int_0^x f(t) \, dt = \int_0^1 f(t) \, dt + \int_1^x f(t) \, dt = 0 + (x - 1) = x - 1, \quad x \geq 1.
\]
Therefore, $g'(x) = 0$ if $x < 1$ and $g'(x) = 1$ if $x > 1$. But $g'(1)$ does not exist.

**Example 5.3.** Let $g(x) = \int_x^b e^{-t^2} \, dt$. Find $g'(x)$.

**Solution:** The function $e^{-t^2}$ is a continuous function everywhere as a composition of two continuous functions, the exponential and power
functions. By the property (5.9), \( g(x) = -\int_b^x e^{-t^2} dt \). Therefore, 
\( g'(x) = -e^{-x^2} \) by (5.19). \(\square\)

This example illustrates the general property:

\[
\frac{d}{dx} \int_x^b f(t) \, dt = -\frac{d}{dx} \int_b^x f(t) \, dt = -f(x)
\]

for a continuous \( f \).

### 33.2. The Definite Integral and Antiderivative.

The following theorem establishes the relation between the definite integral of a function and its antiderivative.

**Theorem 5.6 (The Fundamental Theorem of Calculus).** If \( f \) is continuous on \([a, b]\), then

\[
\int_a^b f(x) \, dx = F(b) - F(a),
\]

where \( F \) is any antiderivative of \( f \), that is, a function such that \( F' = f \).

**Proof.** Let \( g(x) = \int_a^x f(t) \, dt \). By (5.19), the function \( g(x) \) is an antiderivative of \( f(x) \) in an open interval \((a, b)\). If \( F \) is any other antiderivative of \( f \), then \( F \) and \( g \) may differ only by a constant,

\[
F(x) = g(x) + C, \quad a < x < b.
\]

Also, by the definition of \( g(x) \), \( g(a) = 0 \) and \( g(b) = \int_a^b f(t) \, dt \). The function \( g(x) \) is continuous on \([a, b]\) because \( \lim_{x \to a^+} g(x) = g(a) = 0 \) and \( \lim_{x \to b^-} g(x) = g(b) \). Therefore, \( F(x) \) is also continuous on \([a, b]\) (as the sum of two continuous functions). Hence,

\[
F(b) - F(a) = \left( g(b) + C \right) - \left( g(a) + C \right) = g(b) = \int_a^b f(t) \, dt.
\]

The proof is complete. \(\square\)

The following notations are adopted in the fundamental theorem of calculus:

(5.20) \[
\int_a^b f(x) \, dx = F(x) \bigg|_a^b = F(b) - F(a)
\]

The vertical bar at an antiderivative \( F(x) \) means that one has to calculate the difference of the values of \( F \) at the indicated points. The fundamental theorem of calculus provides a powerful analytic tool to evaluate definite integrals.
Example 5.4. Evaluate \( \int_0^1 (1 + x^2)^{-1} \, dx \).

Solution: An antiderivative of \((1+x^2)^{-1}\) is \( F(x) = \tan^{-1} x \). Therefore,
\[
\int_0^1 \frac{1}{1 + x^2} \, dx = \tan^{-1} x \bigg|_0^1 = \tan^{-1}(1) - \tan^{-1}(0) = \frac{\pi}{4} - 0 = \frac{\pi}{4}.
\]

Example 5.5. Evaluate \( \int_1^4 \frac{1 + x}{\sqrt{x}} \, dx \).

Solution: By the linearity of the integral,
\[
\int_1^4 \frac{1 + x}{\sqrt{x}} \, dx = \int_1^4 \left( x^{-1/2} + x^{1/2} \right) \, dx = \int_1^4 x^{-1/2} \, dx + \int_1^4 x^{1/2} \, dx.
\]
An antiderivative of \( x^n \) is \( x^{n+1}/(n+1) \) for any real \( n \neq 1 \). By taking \( n = -1/2 \) and \( n = 1/2 \), an antiderivative is obtained: \( F(x) = 2x^{1/2} + 2x^{3/2}/3 \). Hence,
\[
\int_0^1 \frac{1 + x}{\sqrt{x}} \, dx = \left( 2x^{1/2} + \frac{2x^{3/2}}{3} \right) \bigg|_0^1 = \left( 4 + \frac{16}{3} \right) - \left( 2 + \frac{2}{3} \right) = \frac{20}{3}.
\]

If an object moves along a straight line, its position relative to a fixed point on the line (the origin) may be defined by a single coordinate \( x \) which is a function of time. The velocity \( v(t) = x'(t) \) is positive if the particle moves in the direction in which \( x \) increases and is negative if it moves in the opposite direction. The acceleration is \( a(t) = v'(t) = x''(t) \). A law according to which the acceleration changes with time is usually established by the laws of physics. Then a practical question is to find the position \( x(t) \). Since \( x(t) \) is a second antiderivative of the acceleration, it is not unique and two (initial) conditions must be imposed to get a unique solution.

Example 5.6. A particle moves along the \( x \) axis with the acceleration \( a(t) = 2 - 6t \). Find the position of the particle at the time \( t = 3 \) if its position and velocity at \( t = 1 \) were \( x(1) = 1 \) and \( v(1) = 2 \).

Solution: Since \( v'(t) = a(t) \), the velocity is the antiderivative of the acceleration subject to the condition \( v(1) = 2 \). Hence, by the property (5.10),
\[
v(t) = v(1) + \int_1^t a(s) \, ds = 2 + (2s - 3s^2) \bigg|_1^t = 2t - 3t^2 + 3
\]
Since \( x'(t) = v(t) \), the position \( x(t) \) is the antiderivative of the velocity subject to the condition \( x(1) = 1 \). By property (5.10) such an antiderivative reads

\[
x(t) = x(1) + \int_1^t v(s) \, ds = 1 + (s^2 - s^3 + 3s)
\]

\[
\bigg|_1^t = t^2 - t^3 + 3t - 2
\]

Therefore \( x(3) = -11 \). \( \square \)

### 33.3. Exercises.

1. Find the derivative of each of the following functions:
   
   (i) \( f(x) = \int_1^x (1 + t^6)^{-1} \, dt \)
   
   (ii) \( f(x) = \int_0^x \sin(t^2) \, dt \)
   
   (iii) \( f(x) = \int_{-x}^x \cos(e^t) \, dt \)
   
   (iv) \( f(x) = \int_{\sin x}^x e^{t^2} \, dt \)

2. Let \( f(x) \) be a piecewise constant function:

\[
f(x) = \begin{cases} 
0, & x < 0 \text{ and } x > 6 \\
1, & 0 \leq x < 2 \\
2, & 2 \leq x < 4 \\
-3, & 4 \leq x \leq 6
\end{cases}
\]

Use the geometrical interpretation of the definite integral to draw the graphs of \( g(x) = \int_1^x f(t) \, dt \) and \( h(x) = \int_4^x f(t) \, dt \).

3. For a particle moving down a rough inclined plane, the velocity is \( v(t) = \sqrt{t} \). Where is the particle at the end of 2 seconds?

4. Find the location of a particle moving along a line at the end of 2 seconds if the acceleration of the particle is \( a(t) = 6t - 12t^2 \) and if its position and velocity at the end of 1 second were \( s(1) = 5 \) and \( v(1) = 10 \).

5. A spacecraft had a constant velocity of \( v_0 \). Then its engines were fired for a time \( T_1 \), then stopped for a time \( T \), and then fired again for a time \( T_2 \). If during the time intervals \( T_1 \) and \( T_2 \), the engines created constant accelerations \( a_1 \) and \( a_2 \), respectively, what is the final velocity of the spacecraft?

6. Find the area of the planar region bounded above by the parabola \( y = 3 - x^2 \) and below by the parabola \( y = 1 + x^2 \).

7. Find the area of the planar region bounded above by the parabola \( y = 2 - x^2 \) and below by the line \( y = x \).

8. Evaluate the integrals using the table of antiderivatives:
   
   (i) \( \int_{-1}^5 \sqrt{x} \, dx \)
   
   (ii) \( \int_1^2 (x + 1/x) \, dx \)
(iii) \( \int_{-1/\sqrt{3}}^{\sqrt{3}} (1 + x^2)^{-1} \, dx \)
(iv) \( \int_{-1/2}^{1/2} (1 - x^2)^{-1/2} \, dx \)
(v) \( \int_{-1}^{1} (1 + x^2)^{-1/2} \, dx \)
(vi) \( \int_{0}^{2} |1 - x| \, dx \)
(vii) \( \int_{-\pi/2}^{\pi/2} e^{-2x} \, dx \)
(viii) \( \int_{0}^{\pi} \sin x \, dx \)
(ix) \( \int_{0}^{\pi} \cos^2 x \, dx \)
(x) \( \int_{0}^{1} 6(\sqrt{x} - \sqrt{x}) \, dx \)
(xi) \( \int_{1}^{3} x^2(4 + x^2)^{-1} \, dx \)
(xii) \( \int_{0}^{1} (1 + x + x^2 + x^3 + \cdots + x^n) \, dx \)
(xiii) \( \int_{0}^{2\pi} (1 + \cos x + \cos(2x) + \cdots + \cos(nx)) \, dx \)
(xiv) \( \int_{-1}^{1} (x^2 - 2x \cos \theta + 1)^{-1} \, dx \)

*Hint:* To find an antiderivative, complete the squares.

### 34. Indefinite Integrals and the Net Change

As has been shown in the previous section, the derivative of the definite integral of a continuous function \( f \) with respect to the upper limit equals the value of \( f \) at the upper limit. So integration and differentiation appear as operations inverse to one another. To further stress this relation between the integration and differentiation, the notion of an indefinite integral is introduced.

**Definition 5.6 (Indefinite Integral).** The function \( F \) is called an **indefinite integral** of \( f \) and is denoted by

\[
F(x) = \int f(x) \, dx \quad \text{if} \quad F'(x) = f(x)
\]

It follows from this definition that an indefinite integral is nothing but the general antiderivative of \( f \). The reason for introducing the integral symbol into the antiderivative notation is the fundamental theorem of calculus:

\[
\int_{a}^{b} f(x) \, dx = F(b) - F(a)
\]

where \( F \) is any antiderivative of \( f \). Since all antiderivatives differ only by a constant, which is always cancelled out in the difference \( F(b) - F(a) \), the definite integral is the difference in values of the indefinite integral at the upper and lower limits of the definite integral.
The indefinite integral has the same properties as the antiderivative. It is linear:

\[ \int \left( c_1 f(x) + c_2 g(x) \right) \, dx = c_1 \int f(x) \, dx + c_2 \int g(x) \, dx \]

for any constants \( c_1 \) and \( c_2 \) and any functions \( f \) and \( g \).

Using the table of antiderivatives of basic functions, one can make a table of indefinite integrals of basic functions. Let \( C \) be an arbitrary constant. The following table can be verified by differentiation.

**Table of basic indefinite integrals**

<table>
<thead>
<tr>
<th>Expression</th>
<th>Indefinite Integral</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^n , dx )</td>
<td>( \frac{x^{n+1}}{n+1} + C ), ( n \neq 1 )</td>
</tr>
<tr>
<td>( \sin(ax) , dx )</td>
<td>( -\frac{\cos(ax)}{a} + C )</td>
</tr>
<tr>
<td>( e^x , dx )</td>
<td>( e^x + C )</td>
</tr>
<tr>
<td>( \frac{dx}{1 + x^2} )</td>
<td>( \frac{\tan^{-1} x + C}{2} )</td>
</tr>
<tr>
<td>( \sec^2(ax) , dx )</td>
<td>( \frac{\tan(ax)}{a} + C )</td>
</tr>
<tr>
<td>( \csc^2(ax) , dx )</td>
<td>( -\frac{\cot(ax)}{a} + C )</td>
</tr>
<tr>
<td>( \sec x \tan x , dx )</td>
<td>( \sec x + C )</td>
</tr>
<tr>
<td>( \csc x \cot x , dx )</td>
<td>( -\csc x + C )</td>
</tr>
<tr>
<td>( \frac{dx}{1 - x^2} )</td>
<td>( \frac{1}{2} \ln \left</td>
</tr>
<tr>
<td>( \frac{dx}{x^2 \pm 1} )</td>
<td>( \ln \left</td>
</tr>
</tbody>
</table>

Recall that the general antiderivative on a given interval is obtained from a particular antiderivative by adding an arbitrary constant. This does not hold for a domain being a union of two or more disjoint intervals (review the properties of antiderivatives). So, in the preceding table, the convention is used that the given expressions for indefinite integrals are valid only in a single interval of continuity of the integrand.

**Example 5.7.** Find a general indefinite integral for \( x^{-3} \).

**Solution:** The function \( x^{-3} \) is not defined at \( x = 0 \). So its domain is the union of two disjoint intervals \((-\infty, 0)\) and \((0, \infty)\). By the first equality in the preceding table \( (n = -3) \),

\[ \int x^{-3} \, dx = -\frac{x^{-2}}{2} + C_1, \quad x > 0; \quad \int x^{-3} \, dx = -\frac{x^{-2}}{2} + C_2, \quad x < 0, \]
where \( C_1 \) and \( C_2 \) are arbitrary constants.

**Example 5.8.** Evaluate \( \int_0^1 [3x^2 - x + 4(1 + x^2)^{-1}] \, dx \).

**Solution:** By the linearity of the indefinite integral (5.21), an indefinite integral of the integrand is \( \frac{x^3}{2} - \frac{x^2}{2} + 4 \tan^{-1} x \). An arbitrary constant in the indefinite integral may be omitted here because, as already noted, it is always cancelled out in the definite integral. Therefore,

\[
\int_0^1 \left[ 3x^2 - x + \frac{4}{1 + x^2} \right] \, dx = \left( \frac{x^3}{2} - \frac{x^2}{2} + 4 \tan^{-1} x \right) \bigg|_0^1 = \frac{1}{2} - \pi,
\]

where \( \tan^{-1}(1) = \pi/4 \) has been used.

---

**34.1. The Net Change Theorem.** Put \( f(x) = F'(x) \) in the fundamental theorem of calculus (5.20). The result obtained is known as the net change theorem.

**Theorem 5.7.** The integral of a continuous rate of change is the net change:

\[
\int_a^b F'(x) \, dx = F(b) - F(a).
\]

The continuity of \( F'(x) \) is crucial for the above relation to hold. A formal application of the theorem without checking the continuity may lead to incorrect results. For example, the following line of calculations is false:

\[
\int_{-1}^{1} \frac{d}{dx} \left( -\frac{1}{x} \right) \, dx = -\frac{1}{x} \bigg|_{-1}^{1} = -1 - 1 = -2
\]

Indeed, the derivative \( F'(x) = (-1/x)' = 1/x^2 \) is not defined at \( x = 0 \). It is not possible to assign any numerical value to \( F'(0) \) to make \( F'(x) \) continuous at \( x = 0 \) because \( \lim_{x \to 0} F'(x) = \infty \) which is not a number. Furthermore, \( F'(x) > 0 \) is strictly positive in any interval and, by the property (5.12), so should be the definite integral if it exists. Hence, the above result cannot possibly be true.

The rate \( F'(x) \) may be positive and negative in the interval \([a, b]\) so that the quantity \( y = F(x) \) may increase and decrease. The difference \( F(b) - F(a) \) represents the net change of \( y \) when \( x \) changes from \( a \) to \( b \). The net change vanishes if \( F(b) - F(a) = 0 \). This does not mean that the quantity \( y \) does not change at all, but rather this might mean, for example, that the quantity \( y \) increases from the value \( F(a) \), then, at some \( c \) in \([a, b]\), it begins to decrease, returning to its initial value when \( x = b \) so that its net change vanishes.
An analogy with an object moving along a straight line can be made to illustrate the net change. Let \( x(t) \) be a position function of the object relative to some point on the line. Then \( x'(t) = v(t) \) is its velocity (note that the velocity can be negative so that the object can move back and forth). The net change of the position over the time interval \([t_1, t_2]\) is

\[
\int_{t_1}^{t_2} v(t) \, dt = x(t_2) - x(t_1).
\]

**Example 5.9.** Suppose an object travels along a straight line with a velocity of \( v(t) = 1 - 2t \). Find the net change of its position over the time interval \([0, 1]\) and the total distance traveled by the object over the same time interval.

**Solution:** 1. The indefinite integral of \( v(t) \) is \( x(t) = t - t^2 + C \). So the net change of the object position is

\[
\int_0^1 v(t) \, dt = \int_0^1 x'(t) \, dt = x(1) - x(0) = 0.
\]

2. Note that the velocity changes its sign at \( t = 1/2 \). So, in the interval \([0, 1/2]\), it is positive (i.e., the object moves to the right from its initial position), then the velocity becomes negative in \([1/2, 1]\) (i.e., the object goes back to the initial point). To find the distance traveled by the object, the absolute value \( |v(t)| \) must be integrated over the interval \([0, 1]\). Think of \( |v(t)| \) as the speed shown on the speedometer of your car; it is always non negative regardless of the direction in which the car is moving.

\[
\int_0^1 |1 - 2t| \, dt = \int_0^{1/2} (1 - 2t) - \int_{1/2}^1 (1 - 2t) \, dt = [x(1/2) - x(0)] - [x(1) - x(1/2)] = 1/2,
\]

where the definition \( |v| = v \) if \( v > 0 \) and \( |v| = -v \) if \( v < 0 \) has been used.

Other examples of the net change includes the volume \( V(t) \) of water in a reservoir between two moments of time

\[
\int_{t_1}^{t_2} V'(t) \, dt = V(t_2) - V(t_1),
\]
where \( V'(t) \) is the rate of change of the volume; the net change of the population growth

\[
\int_{t_1}^{t_2} n'(t) \, dt = n(t_2) - n(t_1),
\]

where \( n'(t) \) is the growth rate; the relation between the cost and marginal cost functions:

\[
\int_{t_1}^{t_2} C'(t) \, dt = C(t_2) - C(t_1);
\]

and similarly for many other quantities.

34.2. Exercises.

(1) Find the indefinite integrals. Assume that \( x \) lies in a single interval of continuity of the integrand.

(i) \( \int 6x^5 \, dx \)
(ii) \( \int x^{-3} \, dx \)
(iii) \( \int (x^{2/3} - x^{-2/3}) \, dx \)
(iv) \( \int (x^2 - 4)^2 \, dx \)
(v) \( \int (1 + \sqrt{x})^3 \, dx \)
(vi) \( \int (\sqrt{x} + a)^2 / \sqrt{x} \, dx \)
(vii) \( \int x / \sqrt{x^2 + 1} \, dx \)
(viii) \( \int (1 + x^{-1}) \sqrt{x} \, dx \)
(ix) \( \int (1 + \cos x + \cos(2x)) \, dx \)
(x) \( \int (x + \sin(4x)) \, dx \)
(xi) \( \int (1 + 2x)/\sqrt{1 - x^2} \, dx \)
(xii) \( \int (1 - x + x^3)/x^2 \, dx \)
(xiii) \( \int x^2/(a^2 + x^2) \, dx \)
(xiv) \( \int (x^2 + 3)/(x^2 - 1) \, dx \)
(xv) \( \int x^{-2}\sqrt{x^4 + x^{-4} + 2} \, dx \)
(xvi) \( \int [\sqrt{1 + x^2} + \sqrt{1 - x^2}]/\sqrt{1 - x^4} \, dx \)
(xvii) \( \int (\sqrt{x^2} + 1 - \sqrt{x^2 - 1})/\sqrt{x^4 - 1} \, dx \)
(xviii) \( \int (2x + 3x^2)^2 \, dx \)
(xix) \( \int (2x^2 - 5x^2 - 1)/10^x \, dx \)
(xx) \( \int (e^{3x} + 1)/(e^x + 1) \, dx \)
(xxi) \( \int \sqrt{1 - \sin(2x)} \, dx, \ 0 \leq x \leq \pi \)

*Hint:* Use the fundamental trigonometric identity.

(xxii) \( \int \cot^2 x \, dx \)
(xxiii) \( \int \tan^2 x \, dx \)
(2) Prove that the existence of the indefinite integral of \( f(x) \) implies the existence of the indefinite integral of \( f(ax + b), \ a \neq 0 \), and
\[
\int f(x)\,dx = F(x) + C \implies \int f(ax + b)\,dx = \frac{1}{a}F(ax + b) + C
\]

(3) Use the result of the previous exercise to find
(i) \( \int (x + a)^{-1}\,dx \)
(ii) \( \int (2x - 3)^{10}\,dx \)
(iii) \( \int \sqrt{1 - 3x}\,dx \)
(iv) \( \int (5x - 2)^{5/2}\,dx \)
(v) \( \int (2 + 3x^2)^{-1}\,dx \)
(vi) \( \int (3 - 2x^2)^{-1}\,dx \)
(vii) \( \int (4 - 3x^2)^{-1/2}\,dx \)
(viii) \( \int (2x^2 - 5)^{-1/2}\,dx \)
(ix) \( \int \csc^2(2x + \pi/4)\,dx \)
(x) \( \int (1 + \cos x)^{-1}\,dx \)
(xi) \( \int (1 - \sin x)^{-1}\,dx \)
(xii) \( \int (1 + \sin x)^{-1}\,dx \)

Hint: Use \( \sin x = 2\sin(x/2)\cos(x/2) \) and the fundamental trigonometric identity.

(4) Find a general indefinite integral:
(i) \( \int x^{-1}\,dx \)
(ii) \( \int x^{1/p}\,dx, \ p > 1 \)
(iii) \( \int \sec^2 x\,dx \)
(iv) \( \int (1 - x^2)^{-1}\,dx \)
(v) \( \int (x^2 - 1)^{-1}\,dx \)

(5) Explain why a formal application of the fundamental theorem of calculus with a given antiderivative \( F(x) \) leads to incorrect results if
(i) \( \int_{-1}^{1} \frac{dx}{x}, \quad F(x) = \ln |x| \)
(ii) \( \int_{-1}^{1} \frac{d}{dx} \tan^{-1}\left(\frac{1}{x}\right)\,dx, \quad F(x) = \tan^{-1}\left(\frac{1}{x}\right) \)

(6) Evaluate
\[
\int_{-1}^{1} \frac{d}{dx} \left(\frac{1}{1 + 2^x}\right)\,dx \quad \text{and} \quad \int_{-1}^{1} \frac{d}{dx} \left(\frac{1}{1 + 2^{1/x}}\right)\,dx.
\]
(7) A particle travels with velocity \( v(t) = \sin(t/2) \). Find the net displacement of the particle over the time interval \([0, 2\pi]\) and the distance traveled by the particle.

(8) A bacteria population grows at an exponential rate \( n'(t) = n_0 \gamma e^{\gamma t} \), where \( n_0 \) is the initial population and \( \gamma \) is a constant. If in the time \( T \) the population has doubled, find the constant \( \gamma \). What is the population at \( t = 10T \) as compared to the initial population?

(9) The decay rate of a radioactive element is proportional to the total amount of the element at each moment of time. Find the law of decay of the radium isotope Ra-226 if at an initial moment of time there were \( n_0 \) grams of Ra-226 and in 1600 years its quantity had decreased by two times. If 1 gram of Ra-226 is deposited into a radioactive storage, how much of it will remain in 800 years?

35. The Substitution Rule

35.1. Indefinite Integrals. An indefinite integral of the derivative \( F'(x) \) is the function \( F(x) \) itself, provided \( F'(x) \) is continuous. Let \( u = F(x) \), where \( u \) is a new variable defined as a differentiable function of \( x \). Consider the differential \( du = F'(x) \, dx \). Then the following equalities hold:

\[
\int F'(x) \, dx = F(x) + C = u + C = \int du,
\]

where \( C \) is an arbitrary constant and the last equality follows from the fact that an indefinite integral of \( f(u) = 1 \) is \( u \). So we can conclude that \( \int F'(x) \, dx = \int du \), provided the variables \( u \) and \( x \) are related as \( u = F(x) \). This also shows that it is permissible to operate with \( dx \) and \( du \) after the integral sign as if they were differentials. This observation leads to a neat technical trick to calculate indefinite integrals. For example,

\[
\int \frac{1}{\sqrt{x+1}} \, dx = \int d\left(2\sqrt{x+1}\right) = 2\sqrt{x+1} + C,
\]

where the substitution \( u = 2\sqrt{x+1} \) has been used. This trick can be generalized.

Let \( F(u) \) be an indefinite integral of a continuous function \( f(u) \) on an interval \( I \). Let \( u = g(x) \), where \( g \) is differentiable and its range is the interval \( I \). By the chain rule,

\[
\left(F(g(x))\right)' = F'(g(x))g'(x) = f(g(x))g'(x).
\]
In other words, \( F(g(x)) + C \) is an indefinite integral of \( f(g(x))g'(x) \). On an interval, the most general indefinite integral of \( f(u) \) is \( \int f(u) \, du = F(u) + C \). Therefore, \( F(g(x)) \) and \( \int f(u) \, du \) can differ at most by an additive constant. This proves the following theorem.

**Theorem 5.8.** (The Substitution Rule). If \( u = g(x) \) is a differentiable function whose range is an interval \( I \) and \( f \) is continuous on \( I \), then

\[
\int f(g(x))g'(x) \, dx = \int f(g(x)) \, dg(x) = \int f(u) \, du.
\]

The substitution rule is often referred to as a change of the integration variable. It is a powerful method to calculate indefinite integrals.

**Example 5.10.** Find \( \int x \sin(x^2 + 1) \, dx \).

**Solution:**

\[
\int x \sin(x^2 + 1) \, dx = \int \sin(x^2 + 1) \frac{1}{2} \, d(x^2 + 1) = \frac{1}{2} \int \sin u \, du
\]

\[
= -\frac{1}{2} \cos u + C = -\frac{1}{2} \cos(x^2 + 1) + C,
\]

where the substitution \( u = x^2 + 1 \) has been used.

**Example 5.11.** Find \( \int \tan x \, dx \).

**Solution:**

\[
\int \tan x \, dx = \int \frac{\sin x}{\cos x} \, dx = -\int \frac{d(\cos x)}{\cos x} = -\int \frac{du}{u}
\]

\[
= -\ln |u| + C = -\ln |\cos x| + C = \ln |\sec x| + C,
\]

where the substitution \( u = \cos x \) and the logarithm property \( \ln(1/a) = -\ln a \) have been used. Here it should be noted that the calculations are valid only in a single interval of continuity of \( \tan x \), e.g., \( -\pi/2 < x < \pi/2 \).

A general idea of the substitution is to transform the integral in question to one of the basic integrals given in the table. Sometimes the task can only be accomplished with several substitutions.

**Example 5.12.** Find \( \int (e^x - 1)^{-1/2} \, dx \).

**Solution:** It is suggestive to transform the integrand to a power function with the help of the substitution:

\[
u = e^x - 1 \quad \Rightarrow \quad x = \ln(u + 1) \quad \Rightarrow \quad dx = \frac{du}{u + 1}
\]
Therefore
\[ \int \frac{dx}{\sqrt{e^x - 1}} = \int \frac{1}{\sqrt{u} u + 1} \, du = 2 \int \frac{d\sqrt{u}}{u + 1} = 2 \int \frac{dv}{v^2 + 1} = 2 \tan^{-1} v + C = 2 \tan^{-1} (\sqrt{u}) + C \]
\[ = 2 \tan^{-1}(\sqrt{e^x - 1}) + C \]

where the second substitution \( v = \sqrt{u}, 2dv = du/\sqrt{u} \), has been made. \( \square \)

35.2. Definite Integrals. The substitution rule can be used to evaluate definite integrals by means of the fundamental theorem of calculus.

**Example 5.13.** Evaluate \( \int_0^2 xe^{x^2} \, dx \).

**Solution:** First, find an indefinite integral:
\[ F(x) = \int xe^{x^2} \, dx = \frac{1}{2} \int e^{x^2} \, dx = \frac{1}{2} \int e^u \, du = \frac{1}{2} e^u + C = \frac{1}{2} e^{x^2} + C. \]
where \( u = x^2 \). By the fundamental theorem of calculus,
\[ \int_0^2 xe^{x^2} \, dx = F(2) - F(0) = \frac{1}{2} (e^4 - 1). \] \( \square \)

Note that, when evaluating the integral, the original variable \( x \) has been restored in the indefinite integral in order to apply the fundamental theorem of calculus. The fundamental theorem of calculus can also be applied directly in the new variable \( u \), provided the range of \( u \) is properly changed. Indeed, in the previous example, the answer could have been recovered from the indefinite integral \( \frac{1}{2} e^u + C \) if \( u = x^2 \) ranges from \( 0 = 0^2 \) to \( 4 = 2^2 \) as \( x \) ranges from 0 to 2. This is especially useful when a calculation of a definite integral requires several changes of the integration variable.

**Theorem 5.9.** (The Substitution Rule for Definite Integrals). If \( g' \) is continuous on \([a, b]\) and \( f \) is continuous on the range of \( u = g(x) \), then
\[ \int_a^b f(g(x))g'(x) \, dx = \int_{g(a)}^{g(b)} f(u) \, du. \] (5.23)

**Proof.** Let \( F \) be an antiderivative of \( f \). Then \( F(g(x)) \) is an antiderivative of \((F(g(x)))' = F'(g(x))g'(x) = f(g(x))g'(x) \). By the fundamental theorem of calculus,
\[ \int_a^b f(g(x))g'(x) \, dx = F(g(x)) \bigg|_a^b = F(g(b)) - F(g(a)). \]
On the other hand, since $F(u)$ is an antiderivative of $f(u)$, the fundamental theorem of calculus yields

$$
\int_{g(a)}^{g(b)} f(u) \, du = F(u)\bigg|_{g(a)}^{g(b)} = F(g(b)) - F(g(a)) .
$$

Since the right-hand sides of these equalities coincide, so must their left-hand sides, which implies (5.23). \hfill \Box

**Example 5.14.** Evaluate $\int_1^e \ln x / x \, dx$.

**Solution:** The integrand can be transformed as

$$
\frac{\ln x}{x} \, dx = \ln x \, d\ln x .
$$

So the substitution $u = \ln x$ can be made. The range of the new integration variable $u$ is determined by the range of the old one: $u = 0$ when $x = 1$ and $u = 1$ when $x = e$. Thus,

$$
\int_1^e \frac{\ln x}{x} \, dx = \int_0^1 u \, du = \left. \frac{u^2}{2} \right|_0^1 = \frac{1}{2} . \hfill \Box
$$

### 35.3. Symmetry

The calculation of a definite integral over a symmetric interval can be simplified if the integrand possesses symmetry properties.

**Theorem 5.10.** Suppose $f$ is continuous on a symmetric interval $[-a,a]$. Then

$$
(5.24) \quad \int_{-a}^{a} f(x) \, dx = 2 \int_{0}^{a} f(x) \, dx \quad \text{if } f(-x) = f(x) \ (f \text{ is even}),
$$

$$
(5.25) \quad \int_{-a}^{a} f(x) \, dx = 0 \quad \text{if } f(-x) = -f(x) \ (f \text{ is odd}).
$$

**Proof.** The integral can be split into two integrals:

$$
\int_{-a}^{a} f(x) \, dx = \left( \int_{-a}^{0} + \int_{0}^{a} \right) f(x) \, dx = - \int_{0}^{-a} f(x) \, dx + \int_{0}^{a} f(x) \, dx .
$$

In the first integral on the very right-hand side, the substitution $u = -x$ is made so that $u = 0$ when $x = 0$ and $u = a$ when $x = -a$ and $dx = -du$. Hence,

$$
- \int_{0}^{-a} f(x) \, dx = \int_{0}^{a} f(-u) \, du .
$$
Figure 5.6. Illustration of the property (5.25). A function is odd if \( f(-x) = -f(x) \). Its integral over a symmetric interval \([-a, a]\) vanishes. The area \( A \) under the graph of \( f \) and above the interval \([0, a]\) is the same as the area above the graph of \( f \) and below the interval \([-a, 0]\) because of the skew symmetry of the function and the symmetry of the interval \([-a, a]\) relative to the reflection \( x \rightarrow -x \). By the property depicted in Figure 5.4, the integral of \( f \) over \([-a, a]\) is \( A + (-A) = 0 \).

\[
\int_{-a}^{a} f(x) \, dx = \int_{0}^{a} f(-u) \, du + \int_{0}^{a} f(x) \, dx.
\]

Now, if \( f \) is even, then \( f(-u) = f(u) \) and (5.24) follows. If \( f \) is odd, then \( f(-u) = -f(u) \) and (5.25) follows.

The geometrical interpretation of this theorem is transparent (see Fig 5.6). Suppose \( f(x) \geq 0 \) for \( 0 \leq x \leq a \). The integral \( \int_{0}^{a} f(x) \, dx = A \) is the area under the graph of \( f \) on \([0, a]\). If \( f \) is even, then, by symmetry, the graph of \( f \) on \([-a, 0]\) is obtained from that on \([0, a]\) by a reflection about the \( y \) axis. Therefore, the area \( \int_{-a}^{0} f(x) \, dx \) must coincide with \( A \). If \( f \) is odd, then its graph on \([-a, 0]\) is obtained by the mirror reflection about the origin so that the area \( A \) appears beneath the \( x \) axis. Hence, \( \int_{-a}^{0} f(x) \, dx = -A \).

Example 5.15. Evaluate \( \int_{-\pi}^{\pi} \sin(x^3) \, dx \).
Solution: Unfortunately, an antiderivative of $\sin(x^3)$ cannot be expressed in elementary functions, and the fundamental theorem of calculus cannot be used. One can always evaluate the integral by taking the limit of the sequence of Riemann sums. An alternative solution is due to a simple symmetry argument. Note that $\sin(x^3)$ is an odd function, $\sin((-x)^3) = \sin(-x^3) = -\sin(x^3)$. The integration interval is also symmetric, $[-\pi, \pi]$. Thus, by property (5.25),

$$\int_{-\pi}^{\pi} \sin(x^3) \, dx = 0.$$  \hfill \Box

Remark. In the previous example, take a partition of $[-\pi, \pi]$ by points $x_k = k \Delta x$, $k = -n, -n + 1, \ldots, -1, 0, 1, \ldots, n - 1, n$, where $\Delta x = \pi/n$. Consider the Riemann sum with sample points being the midpoints. They have the property that $x^*_{-k} = -x^*_k$. It is then straightforward to show that the Riemann sum vanishes because $\sin(x^*_{-k}) = \sin((-x^*_k)^3) = -\sin(x^*_k)^3$ for $k = 1, 2, \ldots, n$ (the terms corresponding to negative $x^*_k$ cancel out the terms corresponding to positive $x^*_k$ in the Riemann sum).

35.4. Exercises.

(1) Use the suggested substitution to find the indefinite integrals:

(i) $\int x^3(x^4 + 1)^{1/3} \, dx, \quad u = x^4 + 1$
(ii) $\int \sin(\sqrt{x})/\sqrt{x} \, dx, \quad u = \sqrt{x}$
(iii) $\int \sin x e^\cos x \, dx, \quad u = \cos x$
(iv) $\int x^2 \sqrt{1 - x^2} \, dx, \quad u = \sin x$
(v) $\int (\ln x)^3/x \, dx, \quad u = \ln x$
(vi) $\int (\tan x)^n \sec^2 x \, dx, \quad n \neq -1, \quad u = \tan x$
(vii) $\int (\cot x)^n \csc^2 x \, dx, \quad n \neq -1, \quad u = \cot x$
(viii) $\int (\sin^{-1} x)^2(1 - x^2)^{-1/2} \, dx, \quad u = \sin^{-1} x$
(ix) $\int e^x(e^{2x} + 1)^{-1} \, dx, \quad u = e^x$
(x) $\int e^{2x}(1 + e^x)^{-1} \, dx, \quad u = 1 + e^x$
(xi) $\int \sin(2x)(1 + \cos^2 x)^p \, dx, \quad u = \cos^2 x$
(xii) $\int \sqrt{e^x + 1} \, dx, \quad u = e^x + 1$

Hint: See Example 5.12 to proceed.

(xiii) $\int \sqrt{1 - e^x} \, dx, \quad u = 1 - e^x$

Hint: See Example 5.12 to proceed.

(2) Use a substitution to find the indefinite integrals:

(i) $\int x \sqrt{1 + 2x} \, dx$
(ii) $\int x^2/\sqrt{2 - 3x} \, dx$
(iii) $\int x \sqrt{1 + x} \, dx$
(iv) \( \int e^{-\sqrt{x}}/\sqrt{x} \, dx \)

(v) \( \int x/(x^4 + 2x^2 + 2) \, dx \)

*Hint:* Complete the squares in the denominator.

(vi) \( \int x \tan^{-1} x (1 + x^2)^{-2} \, dx \)

(vii) \( \int \cos^{-1} x / \sqrt{1 - x^2} \, dx \)

(viii) \( \int \sqrt{1 - x^2} \sin^{-1} x \, dx \)

(ix) \( \int (1 + \cos^2 x + \cos^4 x) \sin x \, dx \)

(x) \( \int \sin^3 x / (1 + \cos^2 x) \, dx \)

(xi) \( \int \sqrt{1 + e^x} + \sqrt{1 - e^{-x}} \, dx \)

*Hint:* Transform the integrand to make the difference of perfect squares in the denominator: \( a^2 - b^2 = (a + b)(a - b) \).

(3) Use a change of variables and/or symmetry to evaluate the definite integrals:

(i) \( \int_0^1 x \sqrt{2 + x^2} \, dx \)

(ii) \( \int_0^1 x \tan^{-1} x / (1 + x^2) \, dx \)

(iii) \( \int_0^{\sqrt{\pi}} x \cos(x^2) \, dx \)

(iv) \( \int_{-1}^1 x^3 e^{x^4} \, dx \)

(v) \( \int_{-2}^2 x (e^{x^2} - e^{-x^2}) \, dx \)

(vi) \( \int_{-2}^2 (e^{x^3} - e^{-x^3}) \, dx \)

(vii) \( \int_a^b g(x) \, dx, g(x) = \int_0^x \cos(t^2) \, dt \)

(viii) \( \int_1^2 (2x + 1) \sqrt{x^2 + x + 3} \, dx \)

(ix) \( \int_0^{\pi/2} \sin(2x) \sqrt{1 + \cos^2 x} \, dx \)

(x) \( \int_0^{\pi/4} (\tan x)^p \sec^2 x \, dx, p > 0 \)

(xi) \( \int_{-\pi/6}^{\pi/6} \tan^3(3x) \sin^5(2x) \, dx \)

(xii) \( \int_0^a \sqrt{1 - e^{-x}} \, dx, a > 0 \)

(xiii) \( \int_{-3}^3 x^3 / (1 + x^6) \, dx \)