MAT 167: Calculus I with Analytic Geometry

James Lambers

November 17, 2011
Contents

1 Functions and Limits 7
  1.1 What is Calculus? ........................................... 7
  1.2 Essential Functions and Their Graphs ...................... 12
    1.2.1 Types of Functions ................................. 13
    1.2.2 Transformations of Graphs .......................... 15
  1.3 The Tangent and Velocity Problems ......................... 18
  1.4 The Limit of a Function .................................. 31
    1.4.1 One-Sided Limits ................................... 37
  1.5 Calculating Limits of Functions .......................... 39
  1.6 Continuity .............................................. 51
  1.7 Limits Involving Infinity ................................ 61
    1.7.1 Infinite Limits .................................... 61
    1.7.2 Limits at Infinity .................................. 64

2 Derivatives 73
  2.1 Derivatives and Rates of Change .......................... 73
    2.1.1 Velocities ........................................... 73
    2.1.2 Tangents ............................................. 74
    2.1.3 Other Rates of Change ................................ 76
    2.1.4 The Derivative ...................................... 79
  2.2 The Derivative as a Function ............................... 80
    2.2.1 Notation .............................................. 82
    2.2.2 Differentiability and Continuity ...................... 83
  2.3 Basic Differentiation Formulas ............................. 85
    2.3.1 Derivatives of Trigonometric Functions ............... 88
  2.4 The Product and Quotient Rules ............................. 91
    2.4.1 Proof of the Power Rule (optional) ................... 93
    2.4.2 Differentiation of General Algebraic Functions ...... 94
    2.4.3 Other Trigonometric Functions ....................... 98
CONTENTS

2.4.4 Rates of Change in the Natural and Social Sciences . 102
2.5 The Chain Rule . . . . . . . . . . . . . . . . . . . . . . . . . . 103
2.6 Higher-Order Derivatives . . . . . . . . . . . . . . . . . . . . 113
2.7 Implicit Differentiation . . . . . . . . . . . . . . . . . . . . . 121
2.8 Related Rates . . . . . . . . . . . . . . . . . . . . . . . . . . . 127

3 Inverse Functions 143
3.1 Exponential Functions . . . . . . . . . . . . . . . . . . . . . . 143
  3.1.1 Basic Exponential Graphs . . . . . . . . . . . . . . . . 144
  3.1.2 Additional Exponential Properties . . . . . . . . . . 146
  3.1.3 Applications . . . . . . . . . . . . . . . . . . . . . . . 150
3.2 The Natural Exponential Function . . . . . . . . . . . . . . . 153
  3.2.1 Continuously Compounded Interest . . . . . . . . . . 154
  3.2.2 Rate of Change of the Exponential Function with Base e 155
  3.2.3 Exponential Growth Phenomena . . . . . . . . . . . . 155
3.3 Inverse Functions . . . . . . . . . . . . . . . . . . . . . . . . 157
3.4 General Logarithmic Functions . . . . . . . . . . . . . . . . 163
  3.4.1 Properties of Logarithmic Functions . . . . . . . . . . 164
  3.4.2 Graphs of Logarithmic Functions . . . . . . . . . . . 167
3.5 The Natural and Common Logarithmic Functions . . . . . . . 167
  3.5.1 Definition and Evaluation . . . . . . . . . . . . . . . . 167
  3.5.2 The Change-of-Base Formula . . . . . . . . . . . . . . 170
  3.5.3 Applications . . . . . . . . . . . . . . . . . . . . . . . 171
3.6 Derivatives of Logarithmic and Exponential Functions . . . . 173
  3.6.1 Derivatives of Logarithmic Functions . . . . . . . . . . 174
  3.6.2 Logarithmic Differentiation . . . . . . . . . . . . . . . 176
  3.6.3 Derivatives of Exponential Functions . . . . . . . . . . 177
3.7 Exponential Growth and Decay . . . . . . . . . . . . . . . . 180
  3.7.1 Population Growth . . . . . . . . . . . . . . . . . . . 181
  3.7.2 Radioactive Decay . . . . . . . . . . . . . . . . . . . . 182
  3.7.3 Continuously Compounded Interest . . . . . . . . . . 182
3.8 Inverse Trigonometric Functions . . . . . . . . . . . . . . . . 184
3.9 Indeterminate Forms and l'Hospital's Rule . . . . . . . . . . . 189

4 Applications of Derivatives 197
4.1 Maximum and Minimum Values . . . . . . . . . . . . . . . . 197
4.2 The Mean Value Theorem . . . . . . . . . . . . . . . . . . . . 203
4.3 How Derivatives Affect the Shape of a Graph . . . . . . . . 211
4.4 A Summary of Curve Sketching . . . . . . . . . . . . . . . . 218
4.5 Optimization . . . . . . . . . . . . . . . . . . . . . . . . . . . 228
CONTENTS

4.6 Newton’s Method ........................................... 242
4.7 Antiderivatives ............................................. 248

Index .......................................................... 252
Chapter 1

Functions and Limits

1.1 What is Calculus?

The purpose of mathematics has always been to aid us in describing and relating objects, be they real or conceptual, that have a quantitative nature. We are already familiar with the use of numbers to describe notions such as the length of an object, the amount of time that has elapsed, the area of a surface, or the distance an object has travelled. We have also seen how mathematical principles are used to relate such concepts; for instance, we know that the area of a rectangle, in square inches, is equal to the product of its length and width, provided both quantities are expressed in inches.

Unfortunately, the mathematics with which we are familiar is very limited in terms of its descriptive power, which in turn inhibits our ability to communicate regarding certain concepts and therefore solve problems that involve them. To illustrate such limitations, we consider the concept of velocity. Intuitively, we know what the velocity of an object is: it is the distance that the object travels per unit of time. Furthermore, in the simple case where the object is traveling at constant speed, we know how to compute its velocity, which we will denote by the letter $v$: we simply divide the distance $d$ that it travels by the amount of time $t$ during which it travelled. That is,

$$v = \frac{d}{t}.$$ \hfill (1.1)

**Example 1.1** If you took your car onto the freeway at 2am and used your cruise control while traveling the 156 miles from Irvine, CA to Bakersfield, CA, and it took you two hours to get there, then your velocity was $156/2 = 78$ miles per hour.
Now, if you tried the same thing at 2pm, your car would most certainly not be travelling at constant speed. Your speedometer can give you an accurate indication of your speed at the present time, but what does it mean, mathematically, to say that your speed at this moment is 15 mph? The above formula for velocity, in this case, does little good. If your entire journey to Bakersfield took 3 hours, then you know that your average speed was $156/3 = 52$ miles per hour, but during those three hours, your speedometer indicated speeds ranging anywhere from 0 to 80 mph.

While the formula for velocity cannot be used directly to compute speed at any particular instant, it can be used indirectly, and this usage illustrates what calculus is all about. If your speedometer wasn’t working because it was no longer calibrated correctly, and you wanted to know how fast you were going at exactly 3pm, you could obtain a decent approximation by using your odometer to measure how many miles you travelled from 3:00pm to 3:01pm, and then dividing by one minute, or $1/60$th of an hour. The result is the average speed, in miles per hour, that you travelled from 3:00 to 3:01. If your speed did not change much during that time, then that average speed is likely to be a reasonably accurate estimate of how fast you were going at exactly 3:00.

Realistically, though, significant variation can occur within a whole minute, so it makes sense that a more accurate approximation can be obtained by using the formula $v = d/t$ with a shorter interval of time. If you measured the distance travelled over successively shorter intervals, such as 30 seconds, 10 seconds, 5 seconds, or even 1 second, what you would find is that your various estimates would converge to a particular number. We say that this number is the limit of the average speed as the length of the interval of time approaches zero. To illustrate, consider the following table that lists measurements of the distance travelled since a particular point in time, along with the average speed at which each distance is travelled.

<table>
<thead>
<tr>
<th>Elapsed time (seconds)</th>
<th>Distance (miles)</th>
<th>Average speed (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>1.25</td>
<td>75</td>
</tr>
<tr>
<td>30</td>
<td>0.62</td>
<td>74.4</td>
</tr>
<tr>
<td>10</td>
<td>0.205</td>
<td>73.8</td>
</tr>
<tr>
<td>5</td>
<td>0.102</td>
<td>73.4</td>
</tr>
<tr>
<td>1</td>
<td>0.0202</td>
<td>72.72</td>
</tr>
</tbody>
</table>

It stands to reason that if, somehow, you were able to measure distance travelled during an interval of time that was infinitely short, then you could determine your exact speed at any given instant.
1.1. WHAT IS CALCULUS?

Although such a measurement is not always practical, we have nonetheless gained an understanding of what is the velocity of an object at any particular instant in time: it is the limit of the average velocity over an interval of time that contains that instant, as the length of the interval gets smaller and smaller, eventually approaching zero. Even though, intuitively, the distance travelled during such an interval of time must also approach zero, we find that the ratio of distance to time converges to a number that, intuitively, represents the velocity at a particular instant.

This example illustrates what the subject of this book, differential calculus, is all about: defining the instantaneous rate of change of one quantity (such as distance) with respect to another (such as time) using the simple formula for the average rate of change, in conjunction with the concept of taking a limit of a sequence of numbers that are obtained using that formula. In Section 1.4 we will precisely define the concept of a limit, in order to more precisely define the instantaneous rate of change. These precise definitions will serve as the building blocks for various formulas and theoretical results that we will then use in solving problems concerning rates of change.

The role of differential calculus in solving such problems is to provide a notation that simply describes concepts, such as the instantaneous rate of change, that would be very tedious to describe otherwise. Just as language allows us to communicate effectively by associating concepts, both simple and complex, with words, calculus allows us to work with more complicated mathematical concepts by associating them with various notations. Just as expanding one’s vocabulary facilitates more efficient communication of ideas, the usage of the notation of differential calculus dramatically streamlines the problem-solving process and also enhances our problem-solving capabilities.

Contrary to one’s experience with mathematics prior to learning calculus, mathematics is not simply about using a static collection of formulas and techniques for solving certain problems. Mathematical problems arise in the first place from our need to better understand the world around us in order to function within it. It also consists of the combination of simpler problem-solving techniques with human intuition and experience in order to develop the more powerful techniques that are needed to solve more difficult problems.

Example 1.2 One day, I drove from Phoenix, AZ to Irvine, CA. During the second hour of the trip, my average speed was 78 mph. This was determined using the formula \( v = \frac{d}{t} \), where \( t = 1 \) hour and the value of \( d \) was obtained by observing my odometer at two different times spaced an hour
apart. However, I certainly did not travel at 78 mph during the entire hour, because I encountered a construction zone and was forced to slow down significantly, while driving in excess of 80 mph the rest of the time (after all, it’s the Arizona desert). How would the highway patrol officer posted at the construction site measure my speed as I passed?

The officer’s radar gun can determine my location at various times as I pass. Suppose that it determined my location first at some time \(t_0\), and then, a little later at time \(t_1\), and later still at time \(t_2\), at positions \(p_0\), \(p_1\) and \(p_2\), respectively. Realistically, these times \(t_0\), \(t_1\), and \(t_2\) are spaced tenths or even hundredths of a second apart from one another. We then assume that from time \(t_0\) to time \(t_1\), I traveled a distance \(d_1 = p_1 - p_0\), and then from time \(t_1\) to time \(t_2\), I traveled a distance \(d_2 = p_2 - p_0\). This is illustrated in Figure 1.1.

The software in the radar gun could determine my average speed over the interval of time from \(t_0\) to \(t_2\) using the formula \(v = \frac{d}{t}\), where \(t = t_2 - t_0\) is the elapsed time, and \(d = d_1 + d_2\) is the total distance traveled during that time. Therefore, an approximation of my speed at time \(t_0\) is given by

\[
v = \frac{d_1 + d_2}{t_2 - t_0} = \frac{(p_1 - p_0) + (p_2 - p_1)}{t_2 - t_0} = \frac{p_2 - p_0}{t_2 - t_0}.
\]

Since only a little time has passed between \(t_0\) and \(t_2\), it is reasonable to assume that my speed could not vary much within that time, so I was, approximately, traveling at constant speed. Therefore, the number \(v\) obtained above can be considered a good approximation of my instantaneous speed at time \(t_0\). It is important to keep in mind the following: if an object is traveling at constant speed over some interval in time, then its instantaneous speed at any time within that interval is equal to the object’s average speed, which is given by the formula \(v = \frac{d}{t}\).

It is also reasonable to assume that the smaller the interval of time, the better approximation that average speed is to instantaneous speed, since instantaneous speed has even less opportunity to vary. Therefore, it can be argued that a better approximation of my speed at \(t_0\) is given by the average speed from time \(t_0\) to time \(t_1\), which is simply

\[
v = \frac{d_1}{t_1 - t_0} = \frac{p_1 - p_0}{t_1 - t_0}.
\]

If my position was measured at other points in time even closer to \(t_0\), then the formula \(v = \frac{d}{t}\) would give successively better approximations to my instantaneous speed at \(t_0\). Such a sequence of approximations will converge to
1.1. WHAT IS CALCULUS?

Figure 1.1: Abstract measurements obtained by a radar gun tracking the motion of my car from east to west (right to left). My car’s position \((p_0, p_1\) and \(p_2\)) is obtained at three different times \(t_0, t_1,\) and \(t_2,\) and the distance traveled between these points in time is measured from the position of the car at those times.

A number as the interval of time shrinks to the single instant \(t_0.\) This number, the limit of the average speed as the amount of time elapsed approaches zero, is defined to be my instantaneous speed at \(t_0.\)

In the next section, we will introduce functions into our discussion, thus making it easier to discuss concepts such as position and time. We will then revisit the concept of instantaneous rate of change using functions, and use a function’s graph to show, geometrically, what the instantaneous rate of change of a function is at a particular point. Then, in the following section, we will define precisely what a limit is, which will enable us to compute limits of functions such as those that define the average rate of change. This discussion will allow us to eventually define precisely what the instantaneous
rate of change of a function is.

1.2 Essential Functions and Their Graphs

Before we can begin our exploration of differential calculus, it makes sense to discuss the general context in which it is used. This context consists of a “real-world” problem that needs to be solved. Such a problem could arise from the need to better understand natural phenomena, such as weather systems, planets or even populations of animals. Alternatively, such a problem could pertain to man-made phenomena, such as the stock market, the operation of a piece of equipment, or traffic flow.

The solution of a real-world problem typically proceeds as follows:

1. The problem is described concretely, so that it can be solved. The language of mathematics is often the best tool for this purpose. A mathematical description of a real-world problem is a mathematical problem that is called a model. Often, a model does not describe the underlying real-world problem exactly, because an exact model may be a mathematical problem that cannot be solved. In such cases, it is necessary to make simplifying assumptions, while still creating a model that provides a reasonably accurate description of the real-world problem.

2. The model is then solved using available mathematical techniques.

3. The solution of the model is then interpreted in the context of the underlying real-world problem to obtain some conclusion. Such a conclusion may be a prediction of future behavior or some other insight into the real-world phenomenon that is being studied.

4. Since the model is not exact, it is necessary to validate the conclusion obtained from the model using whatever data is available. If the conclusion happens to be invalid, then the model must be refined so that it better describes the real-world problem, and then the process is repeated.

Example 1.3 Consider the problem of a CHP officer trying to determine whether a passing car is speeding. The officer’s radar gun models the car’s motion mathematically, interpreting the data that it receives to determine the car’s position as a function of time. The resulting mathematical model is the following problem: what is the rate of change of this function with
1.2. ESSENTIAL FUNCTIONS AND THEIR GRAPHS

respect to time? The software in the radar gun solves the problem and computes a number representing this rate of change. The number is interpreted using the appropriate units, such as feet and seconds, resulting in the conclusion that the car is travelling at approximately 90 mph. The officer can validate this conclusion by taking off after the car and comparing its speed to his own. □

In this problem, a key ingredient of the model was a function that described the motion of the car over time. Mathematical models typically include relationships between quantities such as position, velocity, or time, and functions are often ideal for describing such relationships. In this book, we will work with functions that describe how a single quantity, called the dependent variable, depends on the value of a second quantity, called the independent variable. The previous example featured a function in which the dependent variable was the position of the car, and the independent variable was time.

1.2.1 Types of Functions

Because of the essential role that such functions play in mathematical models, we now take some time to review certain types of functions which will be used frequently in this book.

• A linear function is a function whose graph is a straight line. Such a function is usually described using the slope-intercept form

\[ y = mx + b, \tag{1.4} \]

where \( y \) is the dependent variable and \( x \) is the independent variable. The number \( m \) is the slope of the function, and it indicates the rate of change of \( y \) with respect to \( x \). A linear function is the only type of function for which this rate of change is constant; that is, it is independent of \( x \). The number \( b \) is called the y-intercept; it indicates where the line crosses the \( y \)-axis. The equation for a line can be determined by any two points on the line. If \((x_1, y_1)\) and \((x_2, y_2)\) are two points on a line, then the line has the equation

\[ y = m(x - x_1) + y_1 \tag{1.5} \]

where

\[ m = \frac{y_2 - y_1}{x_2 - x_1}. \tag{1.6} \]

It follows that the y-intercept is given by \( b = y_1 - mx_1 \).
Example 1.4 The line passing through the points $(2, 2)$ and $(5, 8)$ has the equation

$$y = \frac{8 - 2}{5 - 2}(x - 2) + 2 = \frac{6}{3}(x - 2) + 2 = 2(x - 2) + 2 = 2x - 4 + 2 = 2x - 2.$$  
(1.7)

A polynomial function is any function of the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0, \quad a_n \neq 0, \quad (1.8)$$

where $n$ is a nonnegative integer that is called the degree of the polynomial. The numbers $a_0, a_1, \ldots, a_n$ are called the coefficients of the polynomial.

A polynomial of degree zero is simply a constant function $y = a_0$. A polynomial of degree one is a linear function. A polynomial of degree 2 is a quadratic function; the graph of such a function is a parabola. A polynomial of degree 3 is a cubic function.

A power function is a function of the form

$$y = x^a$$  
(1.9)

where $a$ is any real number.

Example 1.5 $a = 1/2$ for the square root function $y = \sqrt{x}$. □

A rational function is a function of the form

$$r(x) = \frac{p(x)}{q(x)}$$  
(1.10)

where $p(x)$ and $q(x)$ are both polynomial functions.

Example 1.6 The function $f(x) = 1/x$ is both a power function and a rational function, for it can be written as $f(x) = x^a$ where $a = -1$, and it is also the quotient of the polynomials $p(x) = 1$ and $q(x) = x$. □

An algebraic function is a function that is obtained by applying any number of algebraic operations (addition, subtraction, multiplication, division, and taking roots) to polynomial functions.
Example 1.7  The function
\[ a(x) = \left( \frac{\sqrt{x} - (x^2 + 3x)^3}{1 - \left(1 + \frac{1}{x^4}\right)^{2/5}} \right)^2 + x - 3 \] (1.11)
is an example of an algebraic function. Any polynomial, power or rational function is also an algebraic function. □

- A trigonometric function \( f(t) \) is a function whose value at \( t \) can be obtained by taking any of the possible ratios of the numbers \( x, y, \) and \( r \), where \((x, y)\) is a point on a circle of radius \( r \) centered at the origin \((0, 0)\), and a ray starting at the origin and passing through \((x, y)\) makes an angle of \( t \) radians with the \( x \)-axis.

Example 1.8  Given the definitions of \( x, y \) and \( r \) above, we have the following definitions for the six basic trigonometric functions:
\[ \sin x = \frac{y}{r}, \quad \cos x = \frac{x}{r}, \quad \tan x = \frac{y}{x}, \quad \cot x = \frac{x}{y}, \quad \sec x = \frac{r}{x}, \quad \csc x = \frac{r}{y}. \] (1.12)

If \( 0 < t < \pi/2 \), then these definitions correspond to the right triangle identities that you learned in your trigonometry course. □

1.2.2 Transformations of Graphs

In the previous section, we reviewed various types of functions that will be used throughout this book. From these basic functions, many other functions can be obtained by applying simple transformations, which we now review.

- **Shifting:** adding a constant to either the independent or the dependent variable causes the graph of a function to shift. Adding a positive constant to the independent variable shifts the graph to the left, while adding a negative constant shifts the graph to the right. Adding a positive constant to the dependent variable shifts the graph upward, while adding a negative constant shifts it downward.

Example 1.9  Figure 1.2 illustrates horizontal and vertical shifts of the function \( f(x) = \sin x \). □
Figure 1.2: Top plot: horizontal shifts of $\sin x$. Note that adding a positive constant to $x$ shifts the graph to the left, while using a negative constant shifts to the right. Bottom plot: vertical shifts of $\sin x$. Note that adding a positive constant to the dependent variable shifts the graph up, while a negative constant shifts the graph down.

- **Scaling:** Multiplying the independent or dependent variable by a constant has the effect of scaling the graph in some manner. For instance, multiplying the independent variable by a constant $c$, where $c > 1$, contracts the graph horizontally by a factor of $c$, whereas if $0 < c < 1$, the graph is stretched horizontally by a factor of $1/c$. Similarly, if the dependent variable is multiplied by a constant $c$, where $c > 1$, the graph is stretched vertically by a factor of $c$, whereas it contracts vertically by a factor of $1/c$ if $0 < c < 1$.

**Example 1.10** Figure 1.3 illustrates horizontal and vertical scaling of the function $f(x) = \sin x$. □
1.2. ESSENTIAL FUNCTIONS AND THEIR GRAPHS

Figure 1.3: Top plot: horizontal scaling of $\sin x$. Note that scaling the independent variable by a constant greater than 1 contracts the graph horizontally, while using a positive constant less than 1 stretches it. Bottom plot: vertical scaling of $\sin x$. Scaling the dependent variable by a constant greater than 1 stretches the graph vertically, while a positive constant less than 1 contracts it.

- **Algebraic combinations**: New functions can be obtained by adding, subtracting, multiplying, dividing or exponentiating other functions.

**Example 1.11** The functions $f(x) = x^2$ and $g(x) = \cos x$ can be combined algebraically to obtain the functions $x^2 + \cos x$, $x^2 - \cos x$, $x^2 \cos x$, or $x^2 / \cos x$ (defined wherever $\cos x \neq 0$).

- **Composition**: A new function can be obtained from two functions $f(x)$ and $g(x)$ by the process of composition, in which the function $g$ is applied to the independent variable $x$, and then the function $f$ is
applied to the dependent variable \( g(x) \). This process defines a function \( f \circ g \) whose value at any \( x \) is given by \( f(g(x)) \).

**Example 1.12** The function \( h(x) = \cos(x^2) \) is the composition of the function \( f(x) = \cos x \) and \( g(x) = x^2 \). Specifically, \( h(x) = (f \circ g)(x) \). \( \qed \)

**Example 1.13** Let \( f(x) = \sqrt{x} \) and \( g(x) = x^2 \). Then \((f \circ g)(x) = \sqrt{x^2} = |x| \). On the other hand, \((g \circ f)(x) = (\sqrt{x})^2 = x\), provided that \( x \geq 0 \). \( \Box \)

As illustrated in the previous example, the domain of \((f \circ g)\) is contained within the domain of \( g \). Also, the range of \((f \circ g)\) is contained within the range of \( f \).

### 1.3 The Tangent and Velocity Problems

We now revisit the basic problem of differential calculus: given a function \( f \), how do we compute the instantaneous rate of change of \( y = f(t) \) with respect to \( t \), at a particular point \( t = t_0 \)? For concreteness, one may assume that the independent variable \( t \) denotes time and the dependent variable \( y \) denotes position.

We can approximate this rate of change by choosing some value \( h \) and computing the distance \( d \) traveled during the interval of time from \( t_0 \) to \( t_0 + h \). Then, the average rate of change of \( y = f(t) \) with respect to \( t \) during this interval is given by the ratio of distance to time,

\[
v = \frac{d}{t} = \frac{f(t_0 + h) - f(t_0)}{(t_0 + h) - t_0} = \frac{f(t_0 + h) - f(t_0)}{h}.
\]

(1.14)

Given our interpretations of \( t \) and \( y \), we can think of \( v \) as the average velocity from \( t_0 \) to \( t_0 + h \).

Because \( v \) is equal to the ratio of the difference of two \( y \)-values to the difference in the corresponding \( t \)-values, we can conclude that \( v \) is also equal to the slope of the line that passes through the points \((t_0, f(t_0))\) and \((t_0 + h, f(t_0 + h))\). Because this line passes through two points on the graph of \( f \), we say that the line is a **secant line** of \( f \).

**Definition 1.1** A **secant line** of a function \( f(x) \) is a line that passes through any two distinct points on the graph of \( f \); i.e., it passes through two points of the form \((x_1, f(x_1))\) and \((x_2, f(x_2))\), where \( x_1 \neq x_2 \).
1.3. THE TANGENT AND VELOCITY PROBLEMS

Example 1.14 A secant line for the curve $y = 1/x^6$ is illustrated in Figure 1.4(a). □

Intuitively, we can see that the smaller the value of $h$, the better our approximation of the instantaneous rate of change at $t_0$ by the average rate of change from $t_0$ to $t_0 + h$. As $h$ approaches zero, $t_0 + h$ approaches $t_0$, and it follows that the corresponding secant line converges to a line that intersects the graph of $f$ at the point $(t_0, f(t_0))$, but does not cross the graph at that point. We say that this line is tangent to the graph of $f$ at $t_0$. For completeness we provide a more formal definition of a tangent line.

Definition 1.2 A tangent line of a function $f(x)$ is a line that passes through a point $(x_0, f(x_0))$ on the graph of $f$ and has slope equal to the limit as $h$ approaches 0 of the slope of the secant line of $f$ passing through the points $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$.

Example 1.15 A tangent line for the curve $y = 1/x^6$ is illustrated in Figure 1.4(b). Note that the tangent line intersects the curve at the point (1.2, 1), which is called the point of tangency, the slope of the tangent line can be seen to be equal to the rate of change of $y = 1/x^6$ with respect to $x$. In other words, the direction of the curve and the line appear to be equal at this point. □

Since the average rate of change from $t_0$ to $t_0 + h$ converges to the instantaneous rate of change at $t_0$ as $h$ approaches zero, it follows that the instantaneous rate of change at $t_0$ is equal to the slope of this tangent line.

Example 1.16 Consider the function $f(t) = t^4$, with $t_0 = 1$. Figure 1.5 shows the graph of $f$, along with the secant lines obtained for $h = 1$, 0.5, and 0.1. As $h$ approaches zero, we can see from the figure that the secant line converges to a line that touches the graph of $f$ at the point $(1, 1)$, but does not cross the graph at that point. This line is tangent to the graph of $f$ at the point (1, 1).

To determine the slope of this tangent line, we can compute the slope of the secant line as a function of $h$ and determine the limit of this slope as $h$ approaches zero, as indicated in the above definition of a tangent line. From our previous expression for the average rate of change from $t_0$ to $t_0 + h$, which is equal to the slope of the secant line, we have

$$v = \frac{f(t_0 + h) - f(t_0)}{h}$$
Figure 1.4: (a) Top plot: secant line for the curve $y = 1/x^6$, connecting the points $(1,1)$ and $(1.4,1.4^{-6})$. (b) Bottom plot: tangent line for the same curve. The line is tangent to the graph at the point $(1.2,1.2^{-6})$. Note that the slope of the tangent and secant lines are similar, since the slope of the tangent line is the *instantaneous* rate of change of $y$ with respect to $x$ at $x = 1.2$, while the slope of the secant line is the *average* rate of change of $y$ with respect to $x$ over the interval $1 \leq x \leq 1.4$. 
\[= \frac{(1 + h)^4 - 1}{h} \]
\[= \frac{1 + 4h + 6h^2 + 4h^3 + h^4 - 1}{h} \]
\[= \frac{4h + 6h^2 + 4h^3 + h^4}{h} \]
\[= 4 + 6h + 4h^2 + h^3. \]

As \( h \) approaches zero, we see that the slope of the secant line converges to 4. Therefore, the instantaneous rate of change of \( f(t) \) at \( t = 1 \), which is also the slope of the tangent line at the point \((1, 1)\), is equal to 4. \( \square \)

Figure 1.5: Secant lines of \( f(t) = t^4 \) for various values of \( h \). The slopes of these secant lines represent the average rate of change of \( y \) with respect to \( t \) over the interval from \( t_0 \) to \( t_0 + h \), where \( t_0 = 1 \). Note that as \( h \) approaches 0, the secant line converges to the line that is tangent to the graph of \( f(t) \) at the point \((1, 1)\).
We have demonstrated that the problem of finding the instantaneous rate of change of a function $f$ at a point is equivalent to the problem of finding the slope of the tangent line of $f$ at that point. This geometric interpretation of the rate of change will be a recurring theme throughout this book.

**Example 1.17** Suppose that Wile E. Coyote has, once again, been duped by the Road Runner into heading over a cliff. As usual, he is well past the edge of the cliff before realizing what he has done, at which time he falls straight down. What is his velocity after two seconds of free fall?

**Solution** At time $t$, where $t$ is measured in seconds, his position, or altitude, in feet is given by the function

$$p = -\frac{1}{2}gt^2,$$  \hspace{1cm} (1.15)

where $p = 0$ corresponds to the level of the cliff, and $g$ is the constant of gravity, $32$ ft/s$^2$. Since his altitude is decreasing, $p$ will be negative. Substituting the value of $g$, we have

$$p = -16t^2.$$  \hspace{1cm} (1.16)

This function is shown in Figure 1.6. Note that the graph of this function can be obtained from the graph of $p = t^2$ by first multiplying $t^2$ by $-1$ to reflect its graph across the $p$-axis, and then multiplying it by 16 to stretch the graph vertically by a factor of 16.

His velocity is the instantaneous rate of change at $t = 2$. Since we don’t have a mathematical definition for the velocity at a particular instant (yet), we instead work with what we do know how to compute: the average velocity over an interval in time. Intuitively, our approach to this problem is as follows: we will figure out how to compute the average velocity over an interval in time of the form $[2, 2 + h]$ (that is, the interval $2 \leq t \leq 2 + h$) and find out what happens to the average velocity as the length of the time interval, $h$, approaches zero.

Why does this approach make sense? The reason is that the average velocity over an interval in time is equal to the instantaneous velocity at every point in the interval, if the object is traveling at constant speed. The smaller the interval, the less the object’s speed should vary, so for a very small interval, the average velocity should be a very good approximation to the instantaneous velocity. If the interval is “infinitely small” (i.e., $h = 0$), then we should have the exact value of the instantaneous velocity.

We now recall what is the average velocity on an interval in time: it is the total distance traveled during the interval, divided by the length of the
interval (the elapsed time). In general, the average velocity of an object that has traveled from position \( y_1 \) at time \( t_1 \) to position \( y_2 \) at time \( t_2 \) is
\[
\frac{\text{distance traveled}}{\text{elapsed time}} = \frac{y_2 - y_1}{t_2 - t_1}.
\] (1.17)

In this case, the average velocity over the interval \([2, 2 + h]\) is given by
\[
\frac{-16(2 + h)^2 - 16(2)^2}{(2 + h) - 2}.
\] (1.18)

Ideally, we would like to simply set \( h = 0 \) to get the instantaneous velocity immediately, but we cannot do that in this case because we would be dividing by zero. Instead, we will try to simplify this expression and see if that helps.

We have
\[
\frac{-16(2 + h)^2 - (-16(2)^2)}{2 + h - 2} = \frac{-16(4 + 4h + h^2) + 16(4)}{h}
\]
Therefore, the average velocity of free fall over the interval from \( t = 2 \) to 
\( t = 2 + h \), for any length of time \( h \), is \(-64 - 16h \) ft/s\(^2\). Setting \( h = 0 \), we 
find that the instantaneous velocity at time \( t = 2 \) is \(-64 \) ft/s\(^2\).

If we were to substitute various values for \( h \), we would find that the 
average velocity over the interval \([2, 2 + h]\) converges to the instantaneous 
velocity at time \( t = 2 \). For this reason, we define the instantaneous velocity 
at a given instant to be the limit of the average velocity over an interval 
containing that instant, as the length of the interval (in this example, \( h \)) 
approaches zero. The following table illustrates this convergence of the 
average velocity to the instantaneous velocity.

<table>
<thead>
<tr>
<th>( h )</th>
<th>Avg. vel. (ft/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-96</td>
</tr>
<tr>
<td>1</td>
<td>-80</td>
</tr>
<tr>
<td>0.5</td>
<td>-72</td>
</tr>
<tr>
<td>0.1</td>
<td>-65.6</td>
</tr>
<tr>
<td>0.01</td>
<td>-64.16</td>
</tr>
</tbody>
</table>

We now interpret average velocity and instantaneous velocity geometrically. The average velocity over the interval \([2, 2 + h]\) is the change in \( p \) 
between these two times divided by the change in time. This is equal to the 
slope of the line passing through the points \((2, -64)\) and \((2+h, -16(2+h)^2)\). 
This line is called a secant line for the curve \( p = -16t^2 \). A few such secant 
lines are shown in Figure 1.7. Note that as \( h \) approaches zero, these secant 
lines converge to a line that is tangent to the curve at the point \((2, -64)\); that is, it touches the curve but does not cross it. The instantaneous velocity 
is equal to the slope of this tangent line.

Since we know the slope of the tangent line, and we know a point on the 
line, we can obtain an equation for the line. We use the point-slope form 
\[ p - p_0 = m(t - t_0) \] \hspace{1cm} (1.19)

where \( m \) is the slope and \((t_0, p_0)\) is a point on the line. In this example, 
\( m = -64 \), \( t_0 = 2 \) and \( p_0 = -64 \). Therefore, the equation for the tangent 
line is 
\[ p + 64 = -64(t - 2) \] \hspace{1cm} (1.20)
1.3. THE TANGENT AND VELOCITY PROBLEMS

Figure 1.7: Various secant lines (in red) for the curve $p = -16t^2$, and the tangent line (in black) to which they converge as $h$ approaches 0.

which can be rearranged to obtain an equation in slope-intercept form

$$p = mt + b,$$

or, in this case,

$$p = -64t + 64.$$  \hspace{1cm} (1.22)

Note: You are likely accustomed to seeing the equation of a line written using the letters $x$ and $y$ for the independent variable and dependent variable, respectively. In fact, often these letters will be used for those purposes in this book. However, it is important to keep in mind that one should not get too attached to specific letters being used for specific purposes, because it’s the interpretation of the letter, or other symbol, that is important. Just as you are concerned with the meaning of words as you are reading a book in order to comprehend what is being written, you should keep in mind the meaning of mathematical symbols in order to comprehend their usage.
Example 1.18  My car is traveling along a dark country road at night at 45 mph. All of a sudden, a deer darts out into the road and I slam on my brakes and manage to stop in time and avoid hitting the deer. The following table lists my velocity over time, in mph, where $t = 0$ corresponds to the instant at which I first apply the brakes.

<table>
<thead>
<tr>
<th>$t$ (seconds)</th>
<th>$v$ (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

At what rate am I decelerating after 2 seconds, in m/hr$^2$?

Solution  Just as velocity is the rate of change of distance with respect to time, acceleration is the rate of change of velocity with respect to time. In this case, the acceleration will be negative, since I am slowing down, so we will actually find out how quickly I am decelerating.

First, we need a function that models the given data. There are many ways to construct such a function that are beyond the scope of this book, so we will simply work with one such function:

$$v = 3[(t - 4)^2 - 1], \quad (1.23)$$

where $v$ denotes velocity in mph and $t$ denotes time in seconds. It should be noted that the graph of this function can easily be obtained from the graph of $v = t^2$ by applying the following transformations:

1. First, the graph of $v = t^2$ is shifted by 4 units to the right, which yields the curve $v = (t - 4)^2$.
2. Second, the graph is shifted one unit down, which yields the curve $v = (t - 4)^2 - 1$.
3. Finally, the graph is stretched vertically by a factor of 3, which yields the curve $v = 3[(t - 4)^2 - 1]$.

This sequence of transformations is illustrated in Figure 1.8.

As before, we will compute the average deceleration over the interval in time $[2, 2 + h]$ and find out what happens as the elapsed time, $h$ seconds, approaches zero. This average deceleration is given by

$$\frac{\text{change in velocity}}{\text{change in time}} = \frac{3[(2 + h - 4)^2 - 1] - 3[(2 - 4)^2 - 1]}{2 + h - 2} \quad (1.24)$$
1.3. THE TANGENT AND VELOCITY PROBLEMS

Figure 1.8: Transformation of $v = t^2$ (upper left) by shifting to the right 4 units (upper right), shifting down one unit (lower left), and stretching vertically by a factor of 3 (lower right).

which can be simplified as follows:

$$\frac{3[(2 + h - 4)^2 - 1] - 3[(2 - 4)^2 - 1]}{2 + h - 2} = \frac{3[(h - 2)^2 - 1] - 3[(-2)^2 - 1]}{h}$$

$$= \frac{3[(h^2 - 4h + 4) - 1] - 3[4 - 1]}{h}$$

$$= \frac{3(h^2 - 4h + 3) - 3(3)}{h}$$

$$= \frac{3h^2 - 12h + 9 - 9}{h}$$

$$= \frac{3h^2 - 12h}{h}$$

$$= 3h - 12.$$
Setting $h = 0$, we obtain the instantaneous deceleration at time $t = 2$, which is $-12$ miles per hour per second, since $t$ is measured in seconds. To obtain the deceleration in mi/hr^2, we must multiply this result by the number of seconds in an hour, 3600. Our final answer is, therefore, $-43200$ mi/hr^2. This seems like an incredibly large number, under the circumstances, but keep in mind that this is equivalent to only $-17.6\text{ ft/s}^2$. □

**Example 1.19** Let $u(x)$ be a function that describes the temperature, in degrees Celsius, in a rod that is 4 m long, where $x$ denotes the distance, in meters, between any given point on the rod and its left endpoint, which corresponds to $x = 0$. Specifically,

$$u(x) = \sin\left(\frac{\pi}{4}x\right). \quad (1.25)$$

As we travel along the rod from left to right, what is the instantaneous rate of change of the temperature with respect to distance, at the point that is 1 m from the left endpoint of the rod? Also, what is the equation of the tangent line to $u(x)$ at $x = 1$?

**Solution** The point that is 1 m from the left endpoint of the rod corresponds to $x = 1$. The function that models the temperature in the rod, $u(x) = \sin(\pi x/4)$, is chosen so as to satisfy the condition that the temperature is held fixed at 0° C. The graph of this function can be obtained from the graph of $\sin x$ by stretching it horizontally by a factor of $4/\pi$. This stretching is accomplished by multiplying the independent variable, $x$, by the reciprocal of the factor by which we want to stretch, i.e., $\pi/4$. The graph of $u(x)$ is shown in Figure 1.9.

In order to compute the instantaneous rate of change of temperature with respect to distance at $x = 1$, we proceed as in the previous examples and compute the average rate of change over the interval $[1, 1 + h]$ where $h$ is allowed to vary. We obtain

$$\frac{\text{change in temperature}}{\text{change in distance}} = \frac{\sin(\pi(1 + h)/4) - \sin(\pi/4)}{1 + h - 1}. \quad (1.26)$$

To simplify this, we first use the identity

$$\sin(x + y) = \sin x \cos y + \cos x \sin y \quad (1.27)$$

which yields

$$\frac{\sin(\pi/4) \cos(\pi h/4) + \cos(\pi/4) \sin(\pi h/4) - \sin(\pi/4)}{h} \quad (1.28)$$
Figure 1.9: Graph of \( u(x) = \sin(\pi x/4) \). The red circle marks the point \( x = 1, u = \sqrt{2}/2 \). The red line is tangent to the curve at that point.

or

\[
\sin(\pi/4)[\cos(\pi h/4) - 1] + \cos(\pi/4)\sin(\pi h/4).
\] (1.29)

We then use the half-angle formula

\[
\sin^2 x = \frac{1 - \cos 2x}{2}
\] (1.30)

to obtain

\[
-2\sin(\pi/4)\sin^2(\pi h/8) + \cos(\pi/4)\sin(\pi h/4).
\] (1.31)

Rearranging yields

\[
-2\sin(\pi/4)\sin(\pi h/8)\frac{\sin(\pi h/8)}{h} + \cos(\pi/4)\frac{\sin(\pi h/4)}{h}.
\] (1.32)
We multiply and divide the first term by $\pi/8$, and multiply and divide the second term by $\pi/4$ to obtain

$$-2 \sin(\pi/4) \sin(\pi h/8) \frac{\sin(\pi h/8)}{\pi h/8}\pi + \cos(\pi/4) \sin(\pi h/4) \frac{\pi}{\pi h/4}. \quad (1.33)$$

We now use the fact that as $\theta$ approaches 0, the expression

$$\frac{\sin \theta}{\theta} \quad (1.34)$$

approaches 1. Applying this result with $\theta = \pi h/8$ in the first term, and with $\theta = \pi h/4$ in the second term, we find that as $h$ approaches 0, the average rate of change converges to

$$-2 \sin(\pi/4) \sin(0) \frac{\pi}{8} + \cos(\pi/4) \frac{\pi}{4}, \quad (1.35)$$

and since $\sin(0) = 0$, $\sin(\pi/4) = \sqrt{2}/2$, and $\cos(\pi/4) = \sqrt{2}/2$, the instantaneous rate of change is equal to

$$\left(\frac{\sqrt{2} \pi}{2} \right)^\circ \quad (1.36)$$

To obtain the equation of the tangent line to $u(x)$ at $x = 1$, we need its slope, which is

$$\frac{\sqrt{2} \pi}{8} \approx 0.5554 \quad (1.37)$$

and a point on the line, which is the point $(1, u(1)) = (1, \sin(\pi/4)) = (1, \sqrt{2}/2)$. Using the point-slope form

$$u - u_0 = m(x - x_0) \quad (1.38)$$

with $m = \sqrt{2} \pi/8$, $x_0 = 1$, and $u_0 = \sqrt{2}/2$, we obtain the equation

$$u - \frac{\sqrt{2}}{2} = \frac{\sqrt{2} \pi}{8}(x - 1), \quad (1.39)$$

which can be rearranged to obtain the slope-intercept form of the equation,

$$u = \frac{\sqrt{2} \pi}{8} x + \left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2} \pi}{8}\right). \quad (1.40)$$

This tangent line is shown in Figure 1.9.
1.4 The Limit of a Function

In the previous section we discussed how the average rate of change of a function \( f(x) \) over an interval of the form \([x_0, x_0 + h]\) converges to a number as the interval converges to a single point \( x_0 \), and this number is interpreted as the instantaneous rate of change of the function at the point \( x_0 \). We have referred to this number as the limit of the average rate of change as the width of the interval, \( h \), approaches zero.

If we define the function \( A(h) \) to be the average rate of change of \( f(x) \) over the interval \([x_0, x_0 + h]\), and we denote the instantaneous rate of change of \( f(x) \) at \( x_0 \) by \( A_0 \), then we say that the limit of \( A(h) \), as \( h \) approaches 0, is equal to \( A_0 \). To write this statement concisely, we use the notation

\[
\lim_{h \to 0} A(h) = A_0.
\]

(1.41)

Intuitively, we know what this statement means: as the value of \( h \) gets closer to zero, the value of \( A(h) \) gets closer to \( A_0 \), effectively reaching \( A_0 \) by the time \( h \) reaches zero.

In this section, we define more precisely what it means for a general function \( f(x) \) to have a limit as \( x \) approaches a particular value \( a \). Such a definition will be useful in problems in which it is necessary to compute the limit of a given function.

**Definition 1.3** We write

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

(1.42)

if for any open interval \( I_1 \) containing \( L \), there is some open interval \( I_2 \) containing \( a \) such that \( f(x) \) is in \( I_1 \) whenever \( x \) is in \( I_2 \), and \( x \neq a \). We say that \( L \) is the limit of \( f(x) \) as \( x \) approaches \( a \).

This definition is illustrated in Figure 1.10.

It is not necessarily obvious why this definition of a limit is the correct definition, in the sense that it makes precise the intuitive notion of a limit that we have been using so far. To see why this is the proper definition, let us analyze it carefully. The definition states that we can choose any interval \( I_1 \) containing \( L \), no matter how small, and find some interval \( I_2 \) containing \( a \) such that \( f \) maps any point in \( I_2 \) (except possibly \( a \) itself) to a point in \( I_1 \). Because we can choose the interval \( I_1 \) to be as small as we wish, and because \( f(x) \) belongs to \( I_1 \) for any \( x \) in \( I_2 \), it follows that we can always find an \( x \) near \( a \) such that \( f(x) \) is as close to \( L \) as we wish. Under these
Figure 1.10: Illustration of the definition of a limit, with $f(x) = x^2$ and $a = 0.5$. We can ensure that $0.16 < f(x) < 0.36$ by choosing $x$ so that $0.4 < x < 0.6$. Similarly, we can ensure that $f(x)$ falls within the smaller interval $0.2025 < f(x) < 0.3025$ by requiring that $0.45 < x < 0.55$. In general, for any interval $I_1$ containing 0.25, no matter how small, we can find an interval $I_2$ containing 0.5 such that $f(x)$ is in $I_1$ whenever $x$ is in $I_2$. It follows that $\lim_{x \to 0.5} f(x) = 0.25$. 
conditions, it is impossible for \( f(x) \) to approach any value other than \( L \) as \( x \) approaches \( a \).

We can make the definition a little more concrete by imposing sizes on the intervals \( I_1 \) and \( I_2 \), as long as the interval \( I_1 \) can still be of arbitrary size. It can be shown that the following definition is equivalent to the previous one.

**Definition 1.4** We write

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

(1.43)

if, for any \( \epsilon > 0 \), there exists a number \( \delta > 0 \) such that \( |f(x) - L| < \epsilon \) whenever \( 0 < |x - a| < \delta \).

**Example 1.20** Let \( f(x) = x^3 \), and let \( a = 0 \). We choose the interval \( I_1 \) to be \((-a, b)\) where \( a, b > 0 \); that is, \( I_1 \) contains 0. We then define \( \epsilon \) to be the minimum of \( a \) and \( b \). If we choose \( I_2 \) to be the interval \((-\epsilon^{1/3}, \epsilon^{1/3})\), then, for any \( x \) in \( I_2 \), we must have \( f(x) \) in \( I_1 \). To see this, note that \(-\epsilon^{1/3} < x < \epsilon^{1/3}\) implies \(-\epsilon < x^3 < \epsilon\). From our definition of \( \epsilon \), we have \( b \geq \epsilon \) and \(-a \leq -\epsilon \), so we have \(-a < x^3 < b \), and therefore \( f(x) = x^3 \) is in \( I_1 \). Because this is true for any interval \( I_1 \) containing 0, we can conclude from the definition of a limit that

\[
\lim_{x \to 0} x^3 = 0.
\]  

(1.44)

\( \Box \)

**Example 1.21** Consider the function \( f(x) = \sin(1/x) \), where \( x \neq 0 \). We attempt to determine \( \lim_{x \to 0} f(x) \). From the behavior of \( \sin x \), we know that \(-1 \leq f(x) \leq 1\) for all \( x \), so if the limit exists, it must be a value between \(-1\) and \( 1 \). Unfortunately, as can be seen from the graph of \( f(x) \) in Figure 1.11, \( f(x) \) does not approach any value in this range. More precisely, given any interval containing 0, no matter how small, we can find an \( x \) in the interval for which \( f(x) = L \), for any \( L \) between \(-1\) and \( 1 \). Using the notation of the definition of a limit, no suitable interval \( I_2 \) can be found. Therefore, we say that the limit of \( f(x) \) as \( x \) approaches 0 does not exist. \( \Box \)

It is important to note that the definition of a limit specifically excludes consideration of the behavior of \( f(x) \) at \( x = a \). This is important because a limit is meant to describe the behavior of a function near a point, not at a point. As the preceding example illustrates, the function may not even be defined at that point, or it may have a value at that point that is inconsistent with the behavior of the function near the point.
Example 1.22 Given

\[ f(x) = \frac{x^2 - 1}{x - 1}, \quad x \neq 1, \]  

find \( \lim_{x \to 1} f(x) \), if it exists.

Solution The graph of this function is shown in Figure 1.12. The circle at the point (1, 2) is due to \( f(x) \) being undefined at \( x = 1 \). It can be determined from the graph, or by evaluating \( f(x) \) at various \( x \)-values close to 1, that \( f(x) \) clearly approaches 2 as \( x \) approaches 1. A more conclusive way to see that 2 is the limit is to observe that

\[ \frac{x^2 - 1}{x - 1} = \frac{(x + 1)(x - 1)}{x - 1} = x + 1, \]  

and note that as \( x \) approaches 1, \( x + 1 \) must approach 2.
Figure 1.12: Graph of $y = (x^2 - 1)/(x - 1)$.

Note: while it also would have worked to have simply substituted $x = 1$ into $x + 1$, this would not work in all cases, so that strategy should not be used for computing limits in general. Limits are values that functions approach, not values that they attain. The next example illustrates this distinction. $\Box$

**Example 1.23** Given

$$f(x) = \begin{cases} 
\frac{x^2-1}{x-1} & x \neq 1 \\
1 & x = 1
\end{cases}$$

find $\lim_{x\to 1} f(x)$, if it exists.

**Solution** The graph of this function is shown in Figure 1.13. Since the value of $f(x)$ at $x = 1$ does not influence $\lim_{x\to 1} f(x)$, we have, as before,

$$\lim_{x\to 1} f(x) = 2.$$  \hspace{1cm} (1.48)
This example emphasizes that the limit as $x$ approaches 1 depends only on the behavior of $f(x)$ near $x = 1$, not at $x = 1$. Recall that the definition of a limit states that

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

if, for any open interval $I_1$ containing $L$, there is some open interval $I_2$ containing $a$ such that $f(x)$ is in $I_1$ for any $x$ in $I_2$, where $x \neq a$. □

**Example 1.24** Find

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

**Solution** This function is shown in Figure 1.14. This limit is actually quite difficult to determine using the mathematics with which we are familiar, but it can be seen from the graph, or by evaluating the function at various
1.4. THE LIMIT OF A FUNCTION

In some cases, a function $f(x)$ may approach two different values as $x$ approaches $a$: one value as $x$ approaches $a$ from the left (i.e., only considering $x < a$) and the other value as $x$ approaches $a$ from the right (i.e., considering only $x > a$).

**Example 1.25** The Heaviside function $H(t)$, defined by

$$H(t) = \begin{cases} 
0 & t < 0 \\
1 & t \geq 0
\end{cases}$$

(1.52)
is shown in Figure 1.15. As \( t \) approaches 0 from the left, \( H(t) \) clearly approaches 0, since it is equal to 0 for all \( t < 0 \). However, as \( t \) approaches 0 from the right, \( H(t) \) approaches 1, since it is equal to 1 for all \( t \geq 0 \).

Figure 1.15: Heaviside function defined by \( H(t) = 0 \) for \( t < 0 \), and \( H(t) = 1 \) for \( t \geq 0 \)

If a function \( f(x) \) approaches two different values as \( x \) approaches \( a \), one for each direction of the approach, we cannot say it has a limit as \( x \) approaches \( a \), since the definition of a limit requires that \( f(x) \) be arbitrarily close to a single value for \( x \) within an open interval containing \( a \). However, as we shall see, it is still of interest to know if \( f(x) \) approaches some value from a given direction, so we introduce the following definitions.

**Definition 1.5** We write

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

(1.53)
if, for any open interval $I_1$ containing $L$, there is an open interval $I_2$ of the form $(c, a)$, where $c < a$, such that $f(x)$ is in $I_1$ whenever $x$ is in $I_2$. We say that $L$ is the **limit of $f(x)$ as $x$ approaches $a$ from the left**, or the **left-hand limit of $f(x)$ as $x$ approaches $a$**.

Similarly, we write

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

if, for any open interval $I_1$ containing $L$, there is an open interval $I_2$ of the form $(a, c)$, where $c > a$, such that $f(x)$ is in $I_1$ whenever $x$ is in $I_2$. We say that $L$ is the **limit of $f(x)$ as $x$ approaches $a$ from the right**, or the **right-hand limit of $f(x)$ as $x$ approaches $a$**.

As we have stated, a function $f(x)$ has a limit as $x$ approaches $a$ only if $f(x)$ approaches the **same** value as $x$ approaches $a$ from either direction. Conversely, if $f(x)$ does have a limit as $x$ approaches $a$, then certainly the left-hand and right-hand limits exist, and have the same value. In other words,

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

if and only if

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

1.5 Calculating Limits of Functions

Now that we have precisely defined what the limit of a function is, we turn our attention to actually computing limits of certain functions. While we can always resort to the definition to compute the limit of a given function $f(x)$ as $x$ approaches a given value $a$, this is not the most practical course of action. Instead, it is best to use the definition to establish some **laws** that show how limits of more complicated functions can be obtained in terms of limits of simpler functions. Then, after learning how to compute limits for some very simple functions using the definition, we can use the laws to handle more complicated functions.

We begin by learning how to compute limits of the simplest functions of all: the **constant function** $f(x) = c$, where $c$ is any constant, and the **identity function** $f(x) = x$. For the constant function $f(x) = c$, the limit of $f(x)$ as $x$ approaches $a$ must be equal to $c$. Since $f(x) = c$ for all $x$, it is impossible for $f$ to approach any other value. In summary,

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

(1.57)
Example 1.26 Consider the constant function \( f(x) = 4 \), and suppose that we wish to compute its limit as \( x \) approaches 3. For any open interval \( I_1 \) containing 4, we can easily find an interval \( I_2 \) containing 3 such that \( f(x) \) is in \( I_1 \) for \( x \) in \( I_2 \), because no matter what interval \( I_2 \) we choose, \( f(x) = 4 \) on \( I_2 \), and therefore \( f(x) \) is in \( I_1 \). It follows from the definition of a limit that \( f(x) \) must approach 4.

As for \( f(x) = x \), as \( x \) approaches \( a \), \( f(x) \) must also approach \( a \), since \( f(x) = x \). That is, \[ \lim_{x \to a} x = a. \tag{1.58} \]

**Proof** Let \( I_1 \) be any interval containing \( a \). According to the definition, \( \lim_{x \to a} x = a \) if there is some interval \( I_2 \) containing \( a \) such that \( x \) is in \( I_1 \) whenever \( x \) is in \( I_2 \). We can simply choose \( I_1 \) and \( I_2 \) to be the same interval, since they are both required to contain \( a \). \( \Box \)

We now state some of the limit laws. These laws can be proven using the definition of a limit, but we do not concern ourselves with that in this book. These laws show how the problem of computing the limit of a function with a complicated formula can be reduced to the problem of computing limits of simpler functions. We have

\[ \lim_{x \to a} [f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x), \tag{1.59} \]

\[ \lim_{x \to a} [f(x) - g(x)] = \lim_{x \to a} f(x) - \lim_{x \to a} g(x), \tag{1.60} \]

\[ \lim_{x \to a} [cf(x)] = c \lim_{x \to a} f(x), \tag{1.61} \]

\[ \lim_{x \to a} [f(x)g(x)] = \lim_{x \to a} f(x) \cdot \lim_{x \to a} g(x), \tag{1.62} \]

and

\[ \lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}, \quad \lim_{x \to a} g(x) \neq 0. \tag{1.63} \]

Example 1.27 From the previous discussion, we have

\[ \lim_{x \to 2} x = 2, \quad \lim_{x \to 2} 3 = 3. \tag{1.64} \]

Using the above limit laws, we can compute the following limits:

\[ \lim_{x \to 2} x + 3 = \lim_{x \to 2} x + \lim_{x \to 2} 3 = 2 + 3 = 5, \tag{1.65} \]
1.5. CALCULATING LIMITS OF FUNCTIONS

\[
\lim_{x \to 2} x^2 + 3x = \lim_{x \to 2} x(x + 3) = \lim_{x \to 2} x \cdot \lim_{x \to 2} x + 3 = 2(5) = 10, \quad (1.66)
\]

\[
\lim_{x \to 2} x^2 + 3x - 3 = \lim_{x \to 2} x^2 + 3x - \lim_{x \to 2} 3 = 10 - 3 = 7, \quad (1.67)
\]

\[
\lim_{x \to 2} 4x^2 + 12x = \lim_{x \to 2} 4(x^2 + 3x) = 4 \lim_{x \to 2} x^2 + 3x = 4(10) = 40, \quad (1.68)
\]

and

\[
\lim_{x \to 2} \frac{x}{x+3} = \frac{\lim_{x \to 2} x}{\lim_{x \to 2} x + 3} = \frac{2}{5}. \quad (1.69)
\]

\[\square\]

**Example 1.28** In this example, we use a combination of several of the limit laws:

\[
\lim_{x \to 2} \frac{x^2 - 3x}{x + 3} = \frac{\lim_{x \to 2} x^2 - 3x}{\lim_{x \to 2} x + 3} = \frac{5}{\lim_{x \to 2} x^2 - 3x} = \frac{1}{5} \left[ \lim_{x \to 2} x^2 - \lim_{x \to 2} 3x \right] = \frac{1}{5} \left[ \lim_{x \to 2} x \cdot x - 3 \lim_{x \to 2} x \right] = \frac{1}{5} \left[ \lim_{x \to 2} x \cdot \lim_{x \to 2} x - 3(2) \right] = \frac{2}{5} \cdot 2 - 3(2) = \frac{2}{5} \cdot 4 - 6 = -\frac{2}{5}.
\]

\[\square\]

A direct consequence of the limit law for products is the limit law for exponentiation,

\[
\lim_{x \to a} [f(x)]^n = \left[ \lim_{x \to a} f(x) \right]^n, \quad (1.70)
\]

where \(n\) is a positive integer. Applying this law to the function \(f(x) = x\), we obtain

\[
\lim_{x \to a} x^n = a^n. \quad (1.71)
\]

This result, in combination with the preceding limit laws, can be used to prove the following result concerning limits of polynomial functions or rational functions (recall that every polynomial function is also a rational function).
Theorem 1.1 (Direct Substitution Property, Rational Functions) If $f$ is a rational function and $f(x)$ is defined at $x = a$, then

$$
\lim_{x \to a} f(x) = f(a). \quad (1.72)
$$

In other words, we can compute the limit of a rational function $f(x)$ as $x$ approaches $a$ by simply substituting $x = a$ into the formula for $f(x)$.

Example 1.29 Let $f(x) = x^3 + 3x + 1$. Using the limit laws, we have

$$
\lim_{x \to 3} f(x) = \lim_{x \to 3} x^3 + 3x + 1
= \left[ \lim_{x \to 3} x \right]^3 + 3 \lim_{x \to 3} x + \lim_{x \to 3} 1
= 3^3 + 3(3) + 1
= 34.
$$

Alternatively, we can reach the same conclusion by computing $f(3) = 3^3 + 3(3) + 1 = 34$ and applying the Direct Substitution Property.

Example 1.30 Let

$$
f(x) = \frac{x^3 + 2}{x^2 + 1}. \quad (1.73)
$$

To compute $\lim_{x \to -1} f(x)$, we use the Direct Substitution Property to obtain

$$
\lim_{x \to -1} f(x) = f(-1) = \frac{(-1)^3 + 2}{(-1)^2 + 1} = \frac{-1 + 2}{1 + 1} = \frac{1}{2}. \quad (1.74)
$$

Using the definition of a limit, we can compute the limit of the root of a function. We have

$$
\lim_{x \to a} \sqrt[n]{x} = \sqrt[n]{a}, \quad (1.75)
$$

where $n$ is a positive integer and $a$ is assumed to be positive if $n$ is even. This rule can be generalized to obtain the limit of the root of a general function:

$$
\lim_{x \to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \to a} f(x)}, \quad (1.76)
$$

where $n$ is a positive integer and $\lim_{x \to a} f(x) > 0$ if $n$ is even. This rule can be used to extend the Direct Substitution Property to a larger class of functions.
1.5. CALCULATING LIMITS OF FUNCTIONS

**Theorem 1.2** (Direct Substitution Property, Algebraic Functions) If $f$ is an algebraic function and $f(x)$ is defined at $x = a$, then

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

(1.77)

**Example 1.31** Let

$$f(x) = \sqrt{\frac{x^3 + 2}{x^2 + 1}}.
$$

(1.78)

To compute $\lim_{x \to -1} f(x)$, we can use the law for roots to obtain

$$
\lim_{x \to -1} f(x) = \sqrt{\lim_{x \to -1} \frac{x^3 + 2}{x^2 + 1}} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}.
$$

(1.79)

Alternatively, we can use the Direct Substitution Property to obtain

$$
\lim_{x \to -1} f(x) = f(-1) = \sqrt{\frac{(-1)^3 + 2}{(-1)^2 + 1}} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}.
$$

(1.80)

□

In some cases, it may be difficult to compute the limit of a given function $f(x)$ using the limit laws, but we can learn something about the limit by comparing $f$ to simpler function $g$. This leads to the following result.

**Theorem 1.3** If $f(x) \leq g(x)$ for $x$ near $a$, except possibly at $x = a$ itself, and $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ both exist, then

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

(1.81)

A direct consequence of this result is as follows:

**Theorem 1.4** (Squeeze Theorem) If $f(x) \leq g(x) \leq h(x)$ for $x$ near $a$ (except possibly at $x = a$ itself) and

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

(1.82)

then

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

(1.83)

This theorem is called the Squeeze Theorem because the function $g(x)$ is “squeezed” between $f(x)$ and $h(x)$, and since both of these functions approach the same value $L$ as $x$ approaches $a$, $g$ must also approach $L$. 
Example 1.32 Consider the function \( f(x) = x^2 \tan x \). For \(-\pi/4 \leq x \leq \pi/4\), we have \(-1 \leq \tan x \leq 1\). Multiplying through this inequality by \( x^2 \), we find that \(-x^2 \leq f(x) \leq x^2\). Since \( \lim_{x \to 0} x^2 = (\lim_{x \to 0} x)^2 = 0 \), it follows that \( \lim_{x \to 0} -x^2 = 0 \) as well, so we can apply the Squeeze Theorem and conclude that

\[
\lim_{x \to 0} x^2 \tan x = 0.
\] (1.84)

This application of the Squeeze Theorem is illustrated in Figure 1.16.

Example 1.33 Compute

\[
\lim_{x \to 0} x^2 \cos(1/x).
\] (1.85)

Solution For this function, we cannot use limit laws to conclude

\[
\lim_{x \to 0} x^2 \cos(1/x) = \left[ \lim_{x \to 0} x^2 \right] \cdot \left[ \lim_{x \to 0} \cos(1/x) \right]
\] (1.86)
because \( \cos(1/x) \) does not have a limit as \( x \) approaches 0. Like \( \sin(1/x) \), it oscillates more and more rapidly between \(-1\) and \(1\) as \( x \) approaches 0, eventually oscillating “infinitely rapidly.” Limit laws, such as the law that “the limit of a product is equal to the product of the limits”, only apply if all of the limits involved actually exist.

Therefore, we rely on the Squeeze Theorem. Because

\[
-1 \leq \cos(1/x) \leq 1
\]  

for all \( x \neq 0 \), we can multiply through by \( x^2 \) and obtain

\[
-x^2 \leq x^2 \cos(1/x) \leq x^2,
\]

since \( x^2 > 0 \) for all \( x \neq 0 \). Now, since

\[
\lim_{x \to 0} -x^2 = 0 \quad \text{and} \quad \lim_{x \to 0} x^2 = 0,
\]

we can apply the Squeeze Theorem and conclude that

\[
\lim_{x \to 0} x^2 \cos(1/x) = 0.
\]

This has to be true because \( x^2 \cos(1/x) \) is “squeezed” between \(-x^2\) and \(x^2\), two functions that bound \( x^2 \cos(1/x) \) above and below near \( x = 0 \), and have the same limit as \( x \) approaches 0. Those are the two conditions that must be satisfied in order to apply the Squeeze Theorem. \( \square \)

**Example 1.34** Given that

\[
x^3 - 3x^2 + 3x \leq \frac{x^3 - 3x^2 + 5x - 3}{2x - 2} \leq x^2 - 2x + 2,
\]

for \( x \) near 1, compute

\[
\lim_{x \to 1} \frac{x^3 - 3x^2 + 5x - 3}{2x - 2}.
\]

**Solution** For convenience, let

\[
f(x) = x^3 - 3x^2 + 3x, \quad g(x) = \frac{x^3 - 3x^2 + 5x - 3}{2x - 2}, \quad h(x) = x^2 - 2x + 2.
\]

Then we are given that

\[
f(x) \leq g(x) \leq h(x).
\]
We could compute \( \lim_{x \to 1} g(x) \) directly, but this is not easy. We cannot use the Direct Substitution Property, because \( g \) is not defined at 1. We could try to manipulate \( g \) algebraically and arrive at the limit, but given the above bounds on \( g(x) \), we can instead apply the Squeeze Theorem, since we can use the Direct Substitution Property to obtain

\[
\lim_{x \to 1} f(x) = \lim_{x \to 1} x^3 - 3x^2 + 3x = 1^3 - 3(1^2) + 3(1) = 1 - 3 + 3 = 1, \quad (1.95)
\]

and

\[
\lim_{x \to 1} h(x) = \lim_{x \to 1} x^2 - 2x + 2 = 1^2 - 2(1) + 2 = 1 - 2 + 2 = 1. \quad (1.96)
\]

Since \( g(x) \) is bounded by two functions that have the same limit as \( x \) approaches 1, we can apply the Squeeze Theorem to conclude that

\[
\lim_{x \to 1} g(x) = 1. \quad (1.97)
\]

The graphs of all three functions are shown in Figure 1.17. \( \Box \)

Because we have established all of these laws of limits, we are able to avoid resorting to the definition of a limit to compute limits of a wide variety of functions. The process of establishing these laws is illustrative of the fact that in mathematics, two general types of problems need to be solved:

- Specific problems, typically those obtained by modeling real-world problems for which mathematics was developed in the first place, and

- General, abstract problems, typically for the purpose of developing more powerful problem-solving techniques.

It is recommended that one be comfortable with solving both types of problems, in order to be able to apply the power of mathematics effectively and efficiently.

**Example 1.35** From Section 1.3, the position \( p \) of Wile E. Coyote as he fell off the cliff was described by the function

\[
p = -16t^2, \quad (1.98)
\]

where \( t \) denotes seconds since he began to fall. Determine the instantaneous rate of change of \( t \) with respect to \( p \) when \( p = -64 \) ft.
1.5.  CALCULATING LIMITS OF FUNCTIONS

Figure 1.17: The function \( g(x) = \frac{x^3 - 3x^2 + 5x - 3}{2x - 2} \) (middle curve) is squeezed between \( f(x) = x^3 - 3x^2 + 3x \) (lower curve) and \( h(x) = x^2 - 2x + 2 \) (upper curve), so all three functions approach 1 as \( x \) approaches 1.

**Solution** First, we express \( t \) as a function of \( p \). Dividing through \( p = -16t^2 \) by \(-16\), we have
\[
\frac{-p}{16} = t^2, \tag{1.99}
\]
which yields
\[
t = \sqrt{\frac{-p}{16}}. \tag{1.100}
\]
Now, we can compute the instantaneous rate of change of \( t \) with respect to \( p \) at \( p = -64 \). This is accomplished by computing the average rate of change of \( t \) over the interval \(-64 \leq p \leq -64 + h\), which is given by
\[
\frac{\text{change in } t}{\text{change in } p} = \frac{\sqrt{\frac{-(-64+h)}{16}} - \sqrt{\frac{-(-64)}{16}}}{(-64 + h) - 64}, \tag{1.101}
\]
and then observing what happens as the interval \([-64 + h, -64]\) shrinks to a single point \(p = 64\). In other words, we need to compute
\[
\lim_{h \to 0} \frac{\sqrt{\frac{-(-64 + h)}{16}} - \sqrt{\frac{-(-64)}{16}}}{(-64 + h) - 64}.
\]
(1.102)

We cannot apply the Direct Substitution Property to this algebraic function and simply substitute \(h = 0\), because the function is not defined at \(h = 0\). Instead, we will need to manipulate this function algebraically so that the Direct Substitution Property can be applied.

Simplifying, we have
\[
\lim_{h \to 0} \frac{\sqrt{\frac{64 - h}{16}} - \sqrt{\frac{64}{16}}}{h}.
\]
(1.103)

Next, we factor \(\sqrt{1/16}\) out of both terms in the numerator to obtain
\[
\lim_{h \to 0} \frac{1}{4} \frac{\sqrt{64 - h} - \sqrt{64}}{h},
\]
(1.104)

which can be rewritten using the limit law for products as
\[
\left[ \lim_{h \to 0} \frac{1}{4} \right] \left[ \lim_{h \to 0} \frac{\sqrt{64 - h} - \sqrt{64}}{h} \right],
\]
(1.105)

and then simplified using our knowledge of the limit of a constant function to
\[
\frac{1}{4} \lim_{h \to 0} \frac{\sqrt{64 - h} - \sqrt{64}}{h}.
\]
(1.106)

Now, we multiply the numerator and denominator by the conjugate of the expression \(\sqrt{64 - h} - \sqrt{64}\), which is \(\sqrt{64 - h} + \sqrt{64}\). This yields
\[
\frac{1}{4} \lim_{h \to 0} \frac{\sqrt{64 - h} - \sqrt{64}}{h} = \frac{1}{4} \lim_{h \to 0} \frac{\sqrt{64 - h} - \sqrt{64} \sqrt{64 - h + \sqrt{64}}}{\sqrt{64 - h + \sqrt{64}}},
\]
(1.107)

which simplifies as follows:
\[
\frac{1}{4} \lim_{h \to 0} \frac{\sqrt{64 - h} - \sqrt{64} \sqrt{64 - h + \sqrt{64}}}{h} = \frac{1}{4} \lim_{h \to 0} \frac{(64 - h) - 64}{h(\sqrt{64 - h + \sqrt{64}})}
\]
\[
= \frac{1}{4} \lim_{h \to 0} \frac{-h}{h(\sqrt{64 - h + \sqrt{64}})}
\]
\[
= \frac{1}{4} \lim_{h \to 0} \frac{-1}{\sqrt{64 - h + \sqrt{64}}}
\]
\[
= \frac{1}{4} \lim_{h \to 0} \frac{-1}{\sqrt{64 - h + 8}}.
\]
The function
\[ f(h) = \frac{-1}{\sqrt{64 - h + h}} \] (1.108)
is defined at \( h = 0 \), so we can use the Direct Substitution Property to obtain the limit,
\[ \frac{1}{4} \lim_{h \to 0} \frac{-1}{\sqrt{64 - h + h} + 8} = \frac{1}{4} \frac{-1}{\sqrt{64 - 0 + 8}} = \frac{1}{4} \frac{-1}{8 + 8} = -\frac{1}{64}. \] (1.109)

We conclude that the instantaneous rate of change of \( t \) with respect to \( p \) at \( p = -64 \), which is equal to the limit of the average rate of change over the interval \([ -64, -64 + h] \) as \( h \) approaches 0, is equal to \(-1/64 \) s/ft.

Recall from Section 1.3 that the instantaneous rate of change of \( p \) with respect to \( t \) at \( t = 2 \) is equal to \(-64 \). The rate of change of \( t \) with respect to \( p \), at the same point on the graph of \( p = -16t^2 \), \((t = 2, p = -64)\) is equal to the reciprocal, \(-1/64 \). This makes sense because rate of change corresponds to slope, which is the ratio of the change in one variable to the change in another. In this example, we have reversed the role of the variables \( t \) and \( p \), so the numerator and the denominator in the slope have switched places. □

**Example 1.36** Compute
\[ \lim_{x \to 7} \frac{|x - 7|}{x - 7} \] (1.110)

**Solution** The function \(|x - 7|\) can be rather difficult to work with, because it is not an algebraic function. Therefore, what we will do instead is take into account that
\[ |x - 7| = \begin{cases} x - 7 & x - 7 \geq 0, \\ -x + 7 & x - 7 < 0, \end{cases} \] (1.111)
or
\[ |x - 7| = \begin{cases} x - 7 & x \geq 7, \\ -x + 7 & x < 7, \end{cases} \] (1.112)
and then compute the one-sided limits
\[ \lim_{x \to 7^-} \frac{|x - 7|}{x - 7}, \quad \lim_{x \to 7^+} \frac{|x - 7|}{x - 7}. \] (1.113)
If they both exist and are equal to some value \( L \), then we know that \( L \) is also the limit as \( x \to 7 \).
For the left-hand limit, we are approaching $x = 7$ from the left, so we have $x < 7$. Therefore
\[
\lim_{x \to 7^-} \frac{|x - 7|}{x - 7} = \lim_{x \to 7^-} \frac{-7 + x}{x - 7} = \lim_{x \to 7^-} -1 = -1. \tag{1.114}
\]
Similarly, we have
\[
\lim_{x \to 7^+} \frac{x - 7}{x - 7} = \lim_{x \to 7^+} \frac{x - 7}{x - 7} = \lim_{x \to 7^+} 1 = 1. \tag{1.115}
\]
Since the left-hand and right-hand limits do not agree, we conclude that
\[
\lim_{x \to 7} \frac{|x - 7|}{x - 7} \tag{1.116}
\]
does not exist. \(\Box\)

**Example 1.37** Compute
\[
\lim_{x \to 1} f(x) \tag{1.117}
\]
where
\[
f(x) = \begin{cases} 
(x - 1)^2 & x < 1 \\
x - 1 & x > 1 
\end{cases}. \tag{1.118}
\]
Since $f(x)$ is not an algebraic function and is not even defined at $x = 1$, we cannot use the Direct Substitution Property. Instead, it is best to examine each piece of $f(x)$ individually and compute
\[
\lim_{x \to 1^-} f(x) = \lim_{x \to 1} (x - 1)^2 \tag{1.119}
\]
and
\[
\lim_{x \to 1^+} f(x) = \lim_{x \to 1} (x - 1). \tag{1.120}
\]
For both of these limits, we can use the Direct Substitution Property, which yields
\[
\lim_{x \to 1^-} f(x) = \lim_{x \to 1} (x - 1)^2 = 0, \tag{1.121}
\]
and
\[
\lim_{x \to 1^+} f(x) = \lim_{x \to 1} x - 1 = 0. \tag{1.122}
\]
Since the left-hand and right-hand limits are equal, we can conclude that
\[
\lim_{x \to 1} f(x) = 1. \tag{1.123}
\]
\(\Box\)
Example 1.38 Compute
\[
\lim_{x \to 0} |x|
\]
where \( |x| \) is the greatest integer function: \( |x| \) is equal to the greatest integer that is less than or equal to \( x \).

Solution The greatest integer function has the effect of rounding \( x \) down to the nearest integer. For example,
\[
\]
To compute the limit of this function as \( x \) approaches any integer, we need to consider one-sided limits. As \( x \) approaches 0 from the left, \( |x| = -1 \), since \(-1 \leq x < 0\), so the left-hand limit is equal to \(-1\). As \( x \) approaches 0 from the right, \( |x| = 0 \), since \( 0 \leq x < 1 \), so the right-hand limit is equal to 0. It follows that
\[
\lim_{x \to 0} |x|
\]
does not exist, since the left-hand and right-hand limits are not equal. \( \square \)

1.6 Continuity

In the previous section, we learned that in many cases, the limit of a function \( f(x) \) as \( x \) approached \( a \) could be obtained by simply computing \( f(a) \). Intuitively, this indicates that \( f \) has to have a graph that is one continuous curve, because any “break” or “jump” in the graph at \( x = a \) is caused by \( f \) approaching one value as \( x \) approaches \( a \), only to actually assume a different value at \( a \). This leads to the following precise definition of what it means for a function to be continuous at a given point.

Definition 1.6 (Continuity) We say that a function \( f \) is continuous at \( a \) if
\[
\lim_{x \to a} f(x) = f(a).
\]

Example 1.39 Let \( f(x) = x^2 + 1 \). Because \( f \) is a polynomial, it satisfies the Direct Substitution Property that was introduced in the previous section. In other words, for any \( a \) at which \( f \) is defined,
\[
\lim_{x \to a} f(x) = \lim_{x \to a} x^2 + 1 = \lim_{x \to a} x^2 + \lim_{x \to a} 1 = a^2 + 1 = f(a),
\]
and therefore \( f \) is continuous at \( a \). Since \( a \) is arbitrary, and \( f \) is defined everywhere, it follows that \( f \) is continuous everywhere. \( \square \)
Example 1.40 Let \( f(x) = 1/x \). Being a rational function, it also satisfies the Direct Substitution Property, so it is continuous everywhere that it is defined. However, \( f \) is not defined at \( x = 0 \), and therefore \( f \) is not continuous at 0.

The above discussion also suggests how to define what it means for a function to not be continuous at a given point. Intuitively, we want our definition to include the notion that a function is defined near the point at which it fails to be continuous; that is, the failure to be continuous at a point is an isolated occurrence. This leads to the following definition.

**Definition 1.7** *(Discontinuity)* We say that a function \( f \) is **discontinuous** at \( a \) if \( f \) is defined in an open interval containing \( a \), except possibly at \( a \) itself, and \( f \) is not continuous at \( a \). Alternatively, we say that \( f \) **has a discontinuity** at \( a \).

Example 1.41 Discontinuities can occur for a variety of reasons, as illustrate in Figure 1.18. In Figure 1.18(a), \( f(x) = (x^2-1)/(x-1) \) is approaching the value 2 as \( x \) approaches 1, but \( f \) is not defined at 1, so it cannot be continuous there. In Figure 1.18(b), the function

\[
f(x) = \begin{cases} 
  x + 1 & x \neq 1 \\
  3 & x = 1 
\end{cases} 
\]

(1.129)
is approaching 2 as \( x \) approaches 1, but \( f(1) = 3 \), so again it is not continuous there. In both cases, the discontinuity can be “removed” by simply defining \( f(x) \) to be equal to 2 at \( x = 1 \).

In Figure 1.18(c), the Heaviside function

\[
H(t) = \begin{cases} 
  0 & t < 0 \\
  1 & t \geq 0 
\end{cases}
\]

(1.130)
approaches 0 as \( t \) approaches 0 from the left, while it approaches 1 as \( t \) approaches 0 from the right. In other words, the one-sided limits

\[
\lim_{t \to 0^-} H(t) = 0 \quad \text{and} \quad \lim_{t \to 0^+} H(t) = 1
\]

(1.131)
do not agree. This results in a “jump” in the graph of \( H \). Finally, in Figure 1.18(d), the function \( f(x) = 1/x^2 \) fails to be continuous at \( x = 0 \) because \( f \) is not defined at 0. However, this discontinuity cannot be “removed” by simply defining \( f \) at 0, because \( f \) becomes infinite as \( x \) approaches 0; that is, \( f \) has a **vertical asymptote** at 0.
1.6. CONTINUITY

Figure 1.18: Types of discontinuities. In plot (a), upper left, \( f(x) = \frac{x^2 - 1}{x - 1} \) has a discontinuity at \( x = 1 \) that can be removed by defining \( f(1) = 2 \). In plot (b), upper right, the function \( f \) defined by \( f(x) = x + 1 \) for \( x \neq 1 \), \( f(1) = 3 \) has a discontinuity at \( x = 1 \) that can be removed by redefining \( f(1) \) to be equal to 2. In plot (c), lower left, the Heaviside function has a jump discontinuity in its graph at \( t = 0 \). In plot (d), lower right, the function \( f(x) = \frac{1}{x^2} \) is discontinuous at \( x = 0 \) because of its vertical asymptote.
Based on these examples, the following definitions are used to categorize various types of discontinuities.

**Definition 1.8 (Types of Discontinuities)** Suppose that $f$ has a discontinuity at $a$.

1. The discontinuity is **removable** if $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ both exist and are equal, but are not equal to $f(a)$, or $f$ is not defined at $a$.

2. The discontinuity is called a **jump discontinuity** if $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ both exist but are not equal.

3. The discontinuity is **infinite** if $f$ approaches $\infty$ or $-\infty$ as $x$ approaches $a$ from either direction; that is, $f$ has a vertical asymptote at $a$.

**Example 1.42** As discussed earlier in this section, the function $f(x) = x^2 + 1$ is continuous for all $x$. Now, we consider the function

$$f(x) = \frac{x^3 - x^2 + x - 1}{x - 1}.$$  

This function is a rational function, so it satisfies the Direct Substitution Property for all $x$ at which it is defined. In other words,

$$\lim_{x \to a} f(x) = f(a), \quad a \neq 1,$$  

since $f(x)$ is defined for all $x$ except $x = 1$. It follows that $f$ is continuous at $x$ if $x \neq 1$.

We can graph this function by using polynomial division to obtain

$$\frac{x^3 - x^2 + x - 1}{x - 1} = x^2 + 1,$$  

and then graphing the curve $y = x^2 + 1$, keeping in mind that $f(x)$ is not defined at $x = 1$. The graph of this function is shown in Figure 1.19. The discontinuity at $x = 1$ is **removable**; that is, $f(x)$ can be modified to obtain a new function that is continuous at $x = 1$, simply by defining

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

$\square$
Figure 1.19: Graph of curve $y = (x^3 - x^2 + x - 1)/(x - 1)$. The circle at the point $(1, 2)$ indicates that the $y$ is not defined at $x = 1$; this is an example of a removable discontinuity.

As seen in an earlier example, the Heaviside function $H(t)$ approaches $H(0)$ as $t$ approaches 0 from the right, but not from the left. This is an example of one-sided continuity, which we now define precisely in terms of one-sided limits.

**Definition 1.9 (One-sided Continuity)** We say that a function $f$ is **continuous from the right** at $a$ if

$$
\lim_{x \to a^+} f(x) = f(a). \quad (1.136)
$$

Similarly, we say that $f$ is **continuous from the left** at $a$ if

$$
\lim_{x \to a^-} f(x) = f(a). \quad (1.137)
$$
So far, we have discussed continuity, or lack thereof, at a single point. In describing where a function is continuous, the concept of continuity over an interval is useful, so we define this concept now.

**Definition 1.10** *(Continuity on an Interval)* We say that a function \( f \) is **continuous on the interval** \((a, b)\) if \( f \) is continuous at every point in \((a, b)\). Similarly, we say that \( f \) is continuous on

1. \([a, b)\) if \( f \) is continuous on \((a, b)\), and continuous from the right at \( a \).
2. \((a, b]\) if \( f \) is continuous on \((a, b)\), and continuous from the left at \( b \).
3. \([a, b]\) if \( f \) is continuous on \((a, b)\), continuous from the right at \( a \), and continuous from the left at \( b \).

**Example 1.43** The function \( f(x) = x \) is continuous on the entire real number line, which is the interval \((-\infty, \infty)\). \( \Box \)

**Example 1.44** The Heaviside function is continuous on the interval \((-\infty, 0)\) and on the interval \([0, \infty)\), since it is is continuous at every point except 0, and it is continuous from the right at 0. \( \Box \)

Because continuity is defined using limits, the limit laws introduced in the last section can be used to establish corresponding “continuity laws”. These laws make it very easy to determine where complicated functions are continuous. Just as the limit laws allow one to compute a limit of a complicated function in terms of limits of simpler functions, these continuity laws allow one to decompose such a function into simpler functions and analyze their continuity instead.

**Theorem 1.5** *(Continuity Laws)* If \( f \) and \( g \) are continuous at \( a \) and \( c \) is any constant, then the functions \( f + g \), \( f - g \), \( cf \), and \( fg \) are continuous at \( a \). Furthermore, if \( g(a) \neq 0 \), then \( f/g \) is also continuous at \( a \).

**Example 1.45** Because \( f(x) = x \) and \( g(x) = 3 \) both satisfy the Direct Substitution Property and are defined everywhere, we can conclude that they are continuous everywhere. It follows that the functions \( x + 3 \), \( x - 3 \), \( 3 - x \) and \( 3x \) are also continuous everywhere. Since \( g \) is nonzero, we can also conclude that \( f/g = x/3 \) is continuous everywhere. Finally, \( g/f = 3/x \) is continuous at every point except at \( x = 0 \), since \( 3/x \) is not defined at \( x = 0 \). \( \Box \)
In the previous section, we stated that all algebraic functions (a class of functions that includes all polynomials and rational functions) satisfy the Direct Substitution Property. This leads the following statement regarding the continuity of such functions.

**Theorem 1.6 (Continuity of Algebraic Functions)** If $f$ is an algebraic function, then $f$ is continuous on its entire domain.

**Example 1.46** Consider the function

$$f(x) = \frac{\sqrt{x} + 1}{x - 1}.$$  

(1.138)

This is an algebraic function whose domain is all $x$ except for negative $x$ (due to the square root) and $x = 1$ (due to the denominator). In other words, its domain is the union of the intervals $[0, -1)$ and $(1, \infty)$. On these intervals, $f$ is continuous. $\square$

Although it is not so simple to prove, the trigonometric functions also satisfy the Direct Substitution Property, so we can comment on their continuity as well.

**Theorem 1.7 (Continuity of Trigonometric Functions)** Each of the basic trigonometric functions ($\sin x$, $\cos x$, $\tan x$, $\cot x$, $\sec x$, and $\csc x$) are continuous on their respective domains.

**Example 1.47** The functions $\sin x$ and $\cos x$ are continuous everywhere. On the other hand, $\sec x$ is continuous at all points except for $x = \pi/2 + k\pi$, where $k$ is any integer. This is because $\cos x = 0$ at these points, and therefore $\sec x = 1/\cos x$ has vertical asymptotes at these same points. Because $\sec x$ is defined everywhere else, it is also continuous everywhere else, as can be seen from the fact that $\cos x$ is continuous everywhere, and the law that states that the quotient of two continuous functions (in this case, 1 and $\cos x$) is continuous wherever the function in the denominator is nonzero. $\square$

Previously, we discussed how functions could be composed with one another to obtain new functions. Specifically, given two functions $f(x)$ and $g(x)$, the composition of $f$ and $g$, written as $f \circ g$, is the function whose value at any point $x$ in the domain of $g$ is given by $(f \circ g)(x) = f(g(x))$. We now ask: when is such a composition continuous? We would like to be
able to answer this question using our knowledge of the continuity of \( f \) and \( g \). In order for \( f \circ g \) to be continuous at \( a \), we must have

\[
\lim_{x \to a} f(g(x)) = f(g(a)). \tag{1.139}
\]

Suppose that \( g \) is continuous at \( a \), and let \( b = g(a) \). Furthermore, suppose that \( f \) is continuous at \( b \). Then, as \( y \) approaches \( b \), \( f(y) \) approaches \( f(b) \). Intuitively, it would seem that we could replace \( y \) with \( g(x) \) because \( g \) is continuous at \( a \), and conclude that as \( x \) approaches \( a \), \( y = g(x) \) approaches \( b = g(a) \), and therefore \( f(y) = f(g(x)) \) approaches \( f(b) = f(g(a)) \). This is in fact the case, as we formally state now.

**Theorem 1.8 (Continuity of Compositions)** If \( g \) is continuous at \( a \) and \( f \) is continuous at \( g(a) \), then \( (f \circ g) \) is continuous at \( a \).

**Example 1.48** The function \( h(x) = \sqrt{x+1} \) is continuous for all \( x \) in the interval \([-1, \infty)\). This is because the function \( g(x) = x+1 \) is continuous everywhere, and the function \( f(x) = \sqrt{x} \) is continuous on the interval \([0, \infty)\). Because \( g(x) = x+1 \) lies in the interval \([0, \infty)\) for \( x \) in the interval \([-1, \infty)\), this determines where the composition \( f(g(x)) \) is continuous: it is continuous at every \( x \) in the domain of \( g \) for which \( g \) is continuous at \( x \), and \( f \) is continuous at \( g(x) \).

**Example 1.49** Let \( f(x) = 1/(x-1) \) and \( g(x) = x^2 + 1 \). The composition of \( f \) with \( g \), denoted by \( f \circ g \), is a new function defined by

\[
(f \circ g)(x) = f(g(x)). \tag{1.140}
\]

We determine whether \( (f \circ g)(x) \) is continuous at \( x = 1 \). Although \( f \) is not continuous at \( 1 \), this turns out to be irrelevant in this case, because in order for \( f \circ g \) to be continuous at \( a \), it is only necessary for \( g \) to be continuous at \( a \), and for \( f \) to be continuous at \( g(a) \), not \( a \).

Since \( g \) is continuous at \( 1 \), and \( f \) is continuous at \( g(1) = 2 \), it follows that \( f \circ g \) is continuous at \( 1 \). On the other hand, \( f \circ g \) is not continuous at \( 0 \), because although \( g \) is continuous at \( 0 \), and \( f \) is continuous at \( 0 \), \( f \) is not continuous at \( g(0) = 1 \). This can be seen from the fact that

\[
(f \circ g)(x) = f(g(x)) = \frac{1}{g(x) - 1} = \frac{1}{(x^2 + 1) - 1} = \frac{1}{x^2}, \tag{1.141}
\]

which is not even defined at \( x = 0 \). In many cases, it is not as simple as in this example to determine the continuity of \( f \circ g \) directly, but we can see that the continuity of \( f \circ g \) can be determined from the continuity of \( f \) and \( g \), which are usually simpler functions to work with. \( \square \)
1.6. CONTINUITY

We conclude with a statement about the behavior of a continuous function $f$ on some interval $[a, b]$. The graph of such a function is a continuous curve connecting the points $(a, f(a))$ with $(b, f(b))$. If one were to draw such a graph, their pen would not leave the paper in the process, and therefore it would be impossible to “avoid” any $y$-value between $f(a)$ and $f(b)$. This leads to the following statement about such continuous functions.

**Theorem 1.9 (Intermediate Value Theorem)** Let $f$ be continuous on $[a, b]$. Then, on $(a, b)$, $f$ assumes every value between $f(a)$ and $f(b)$; that is, for any value $y$ between $f(a)$ and $f(b)$, $f(c) = y$ for some $c$ in $(a, b)$.

The Intermediate Value Theorem has a very important application in the problem of finding solutions of a general equation of the form $f(x) = 0$, where $x$ is the solution we wish to compute and $f$ is a given continuous function. Often, methods for solving such an equation try to identify an interval $[a, b]$ where $f(a) > 0$ and $f(b) < 0$, or vice versa. In either case, the Intermediate Value Theorem states that $f$ must assume every value between $f(a)$ and $f(b)$, and since 0 is one such value, it follows that the equation $f(x) = 0$ must have a solution somewhere in the interval $(a, b)$.

**Example 1.50** Compute a solution of the equation $f(x) = 0$, where $f(x) = x^2 - x - 1$.

**Solution** The graph of this function is shown in Figure 1.20. To find a solution, we can first look for two points $a$ and $b$ such that $f(a)$ and $f(b)$ have opposite signs. For example, $f(1) = -1$ and $f(2) = 4 - 2 - 1 = 1$. Since $f$ is continuous everywhere, we can apply the Intermediate Value Theorem and conclude that $f(x) = 0$ for some $x$ in the interval $(1, 2)$.

We can apply the Intermediate Value Theorem repeatedly to narrow down our search for a solution, using a procedure called *bisection*. We first compute the midpoint of the interval $(1, 2)$, which is $(1 + 2)/2 = 3/2$. Since $f(2) > 0$, and $f(3/2) = 9/4 - 3/2 - 1 = -1/4 < 0$, we can conclude that $f(x) = 0$ for some $x$ in the interval $(3/2, 2)$.

Repeating this process on the interval $(3/2, 2)$, we obtain the midpoint, $m = (3/2 + 2)/2 = 7/4$. We then evaluate $f$ at this midpoint to obtain $f(7/4) = 49/16 - 7/4 - 1 = 5/16$. Since $f(3/2) < 0$ and $f(7/4) > 0$, we can apply the Intermediate Value Theorem a third time to conclude that $f(x) = 0$ for some $x$ in the smaller interval $(3/2, 7/4)$.

We can continue this process for as long as we wish, until we are searching for a solution within an interval that is so small that we could choose any point in the interval and have an *approximate solution* that is close enough
Figure 1.20: Graph of \( f(x) = x^2 - x - 1 \). Note that \( f(1.5) \) is negative, while \( f(1.75) \) is positive. Therefore, it follows from the fact that \( f \) is continuous that \( f(x) = 0 \) for some \( x \) between 1.5 and 1.75.

to the exact solution. For example, the interval \((3/2, 7/4)\) has a width of 1/4. If we halved this interval eight more times using bisection, then the resulting interval would have a width of 1/1024. If we only wanted a solution that is correct to two decimal places, then we could choose any point in this small interval and consider it our approximate solution. In this example, the solution that would eventually be obtained using bisection, correct to four decimal places, is approximately 1.6180.

It should be noted that this particular equation, \( x^2 - x - 1 = 0 \), can easily be solved using the quadratic formula. However, there are countless equations of the form \( f(x) = 0 \) for which there is no simple formula for obtaining the solution. For example, there is no formula for the solution of \( f(x) = 0 \) where \( f \) is a polynomial of fourth degree or higher. With such a function, a procedure like bisection must be used to obtain an approximate
Example 1.51 Consider the function \( f(x) = x^3 + x + 1 \). Because it is a polynomial, it is continuous everywhere. We wish to solve the equation \( f(x) = 0 \). We note that \( f(0) = 1 \), while \( f(-1) = -1 \). By the Intermediate Value Theorem, we can conclude that \( f \) assumes every value between \(-1\) and \( 1 \) in the interval \((-1, 0)\), including \( 0 \). Therefore, a solution to the equation \( x^3 + x + 1 = 0 \) must lie in this interval.

We can use bisection to find an approximate solution. We evaluate \( f \) at the midpoint of the interval \((-1, 0)\), which is \(-\frac{1}{2}\), and obtain \( f(-\frac{1}{2}) = \frac{3}{8} \). Since \( f(-\frac{1}{2}) > 0 \) and \( f(-1) < 0 \), we can use the Intermediate Value Theorem again to conclude that a solution lies in the interval \((-1, -\frac{1}{2})\).

We can continue this process for as long as we wish, repeatedly cutting the length of our interval in half until it is so small that any point in the interval can be considered a good approximate solution. The solution is approximately \(-0.6823\). □

1.7 Limits Involving Infinity

1.7.1 Infinite Limits

There are many cases in which the value of a function \( f(x) \) becomes infinite as \( x \) approaches a particular point \( a \). The values \( \infty \) and \(-\infty \) are not specific numbers, and therefore it is not correct to say that \( f(x) \) has a limit as \( x \) approaches \( a \). Nonetheless, we find it convenient to use the notation of limits to describe the function’s behavior in these cases.

Definition 1.11 We write

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

if, for any number \( M > 0 \), there exists an interval \( I \) containing \( a \) such that \( f(x) > M \) whenever \( x \) is in \( I \) and \( x \neq a \). Similarly, we write

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

if, for any number \( M < 0 \), there exists an interval \( I \) containing \( a \) such that \( f(x) < M \) whenever \( x \) is in \( I \) and \( x \neq a \).

Example 1.52 The function \( f(x) = 1/x^2 \) satisfies

\[
\lim_{x \to 0} f(x) = \infty
\]
CHAPTER 1. FUNCTIONS AND LIMITS

since \( f(x) \) can be made arbitrarily large by choosing \( x \) sufficiently close to zero. \( \square \)

We can easily modify these definitions to define the statements

\[
\lim_{x \to a^-} f(x) = \infty, \quad \lim_{x \to a^+} f(x) = \infty
\]

(1.145)

\[
\lim_{x \to a^-} f(x) = -\infty, \quad \lim_{x \to a^+} f(x) = -\infty.
\]

(1.146)

This leads to the following definition that, as we will see in Chapter 4, is very useful in sketching graphs of functions.

**Definition 1.12** The function \( f(x) \) has a **vertical asymptote** at \( x = a \) if at least one of the following statements is true:

\[
\lim_{x \to a^-} f(x) = \infty, \quad \lim_{x \to a^+} f(x) = \infty,
\]

(1.147)

\[
\lim_{x \to a^-} f(x) = -\infty, \quad \lim_{x \to a^+} f(x) = -\infty.
\]

(1.148)

**Example 1.53** The function \( f(x) = 1/x \) has a vertical asymptote at \( x = 0 \), since \( \lim_{x \to 0^-} = -\infty \) and \( \lim_{x \to 0^+} = \infty \). \( \square \)

**Example 1.54** The function \( f(x) = \tan x \) has a vertical asymptote at every \( x \)-value of the form \( x = \pi/2 + k\pi \), where \( k \) is an integer (i.e., \( x = \pi/2, 3\pi/2, 5\pi/2, \ldots \) and \( x = -\pi/2, -3\pi/2, \ldots \)). For each such asymptote, \( \tan x \) approaches \( \infty \) as \( x \) approaches the asymptote from the left, and \( -\infty \) as \( x \) approaches the asymptote from the right. The graph of \( \tan x \) and some of its asymptotes are shown in Figure 1.21. \( \square \)

**Example 1.55** What limits, if any, does the function \( f(x) = 1/x \) have as \( x \) approaches 0?

**Solution** We attempt to compute the left-hand and right-hand limits of \( f(x) \). As \( x \) approaches 0 from the right, \( f(x) \) is positive. The value of \( f \) continues to increase; in fact, for any number \( M \), we can find an interval of the form \( 0 < x < a \) such that \( f(x) > M \) for \( x \) in this interval. From the definition of a right-hand infinite limit, we conclude that

\[
\lim_{x \to 0^+} f(x) = \infty.
\]

(1.149)
As $x$ approaches 0 from the left, $f(x)$ is negative. We use the same approach as with the right-hand limit to conclude

$$\lim_{x \to 0^-} f(x) = -\infty.$$  \hfill (1.150)

Since the one-sided limits are different, it follows that $\lim_{x \to 0} f(x)$ does not exist. We can, however, say that $f(x)$ has a vertical asymptote at $x = 0$. \hfill \square

**Example 1.56** Find any vertical asymptotes of the function

$$f(x) = \begin{cases} 
\frac{1}{x-1} & x > 1 \\
\frac{1}{x} & x \leq 1 
\end{cases}$$ \hfill (1.151)

**Solution** The graph of this function is shown in Figure 1.22. For $x > 1$, $f(x)$ is positive, and it increases toward $\infty$ as $x$ approaches 1 from the right.
It follows that
\[ \lim_{x \to 1^+} f(x) = \infty, \]
which implies that \( f(x) \) has a vertical asymptote at \( x = 1 \). This is the only possible asymptote, since \( f(x) \) has a finite value for all other values of \( x \). \( \square \)

Figure 1.22: Graph of \( f(x) = 1/(x-1) \), \( x > 1 \), and \( f(x) = x \), \( x \leq 1 \). The dashed line shows the function’s vertical asymptote.

### 1.7.2 Limits at Infinity

We have just learned about what it means for a function \( f(x) \) to become infinite as \( x \) approaches a given finite value. Now, we consider the opposite scenario: \( f(x) \) approaching a finite value as \( x \) becomes infinite. As we will see in Chapter 4, knowledge of this behavior of a function can be useful for obtaining its graph.

#### Example 1.57

Consider the function \( f(x) = 1/x \), on the interval \((0, \infty)\). Our goal is to graph this function. We know that it has a vertical asymptote...
at $x = 0$, since
\[
\lim_{x \to 0^+} \frac{1}{x} = \infty.
\]
Furthermore, we know that this function is decreasing on $(0, \infty)$, since $f'(x) = -1/x^2 < 0$. Finally, we know that the graph is concave upward, since $f''(x) = 2/x^3$, which is positive for $x > 0$.

However, this is not enough information to determine how to graph $f(x)$ on $(0, \infty)$. We can evaluate $f(x)$ at a number of points, but it would be helpful to have an idea of how the function behaves as $x$ increases and eventually becomes infinite. In this case, $1/x$ becomes smaller and smaller, approaching zero as $x$ continues to increase. In fact, for any number $\epsilon > 0$, no matter how small, we can find an $x$-value such that $1/x < \epsilon$. In other words, $f(x)$ approaches 0 as $x$ approaches $\infty$. With this information, and the value of $f(x)$ at just a few points, along with our other knowledge of the function’s behavior, we can easily draw an accurate graph, such as the one shown in Figure 1.23. □

In the preceding example, a function $f(x)$ was approaching a particular value as $x$ became infinite. It is natural to use the notation of limits to describe this behavior concisely. We therefore define what it means for a function $f(x)$ to have a limit as $x$ becomes infinite.

**Definition 1.13 (Limit at Infinity)** Let $f(x)$ be a function defined on an interval of the form $(a, \infty)$ for some number $a$. We say that the limit of $f(x)$ as $x$ approaches $\infty$ is equal to $L$, and write
\[
\lim_{x \to \infty} f(x) = L,
\]
if for any open interval $I$ containing $L$, there exists a number $M$ such that $f(x)$ is in $I$ whenever $x > M$.

Similarly, if $f(x)$ is defined on an interval of the form $(-\infty, a)$ for some number $a$, then we say that the limit of $f(x)$ as $x$ approaches $-\infty$ is equal to $L$, and write
\[
\lim_{x \to -\infty} f(x) = L,
\]
if for any open interval $I$ containing $L$, there exists a number $M$ such that $f(x)$ is in $I$ whenever $x < M$.

Previously, we defined a vertical asymptote of $f(x)$ to be an $x$-value $a$ at which $f(x)$ becomes infinite as $x$ approaches $a$. Visually, the graph of $f$
approaches the vertical line $x = a$ as $x$ approaches $a$. Similarly, if $f(x)$ has a finite limit $L$ as $x$ approaches $\infty$ or $-\infty$, then we can see that the graph of $x$ is approaching the horizontal line $y = L$. This leads to the following definition.

**Definition 1.14 (Horizontal Asymptote)** The horizontal line $y = L$ is called a **horizontal asymptote** of a function $f(x)$ if either

$$\lim_{x \to \infty} f(x) = L \quad \text{or} \quad \lim_{x \to -\infty} f(x) = L.$$  

(1.156)

**Example 1.58** The function $f(x) = 1/x$ has a horizontal asymptote at $y = 0$, since

$$\lim_{x \to \infty} \frac{1}{x} = 0, \quad \text{and} \quad \lim_{x \to -\infty} \frac{1}{x} = 0.$$  

(1.157)
1.7. LIMITS INVOLVING INFINITY

If we consider the function \( f(x) = 1/x^2 \), we see that it also has a horizontal asymptote at \( y = 0 \). In fact, this is true for the function \( f(x) = 1/x^r \), for any rational number \( r \), as stated in the following result.

**Theorem 1.10** If \( r > 0 \) is a rational number, then

\[
\lim_{x \to \infty} \frac{1}{x^r} = 0.
\]

Furthermore, if \( x^r \) is defined for all \( x \), then

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

It is natural to ask whether the Limit Laws, established to aid in computing limits of functions when approaching finite values, can be used to compute limits when approaching infinite values. This is in fact the case, with the exception of the laws

\[
\lim_{x \to a} x^n = a^n, \quad \lim_{x \to a} \sqrt[n]{x} = \sqrt[n]{a},
\]

since we cannot exponentiate \( \infty \) or \( -\infty \). We now illustrate the use of these laws in determining horizontal asymptotes.

**Example 1.59** We now use the Limit Laws to compute

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

Using the laws for quotients, sums, differences, constant multiples, and constant functions, as well as the preceding theorem on the limit of \( 1/x^r \), we obtain

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

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

\[
= \lim_{x \to \infty} \left(1 + \frac{3}{x} + \frac{4}{x^2}\right) / \lim_{x \to \infty} \left(2 - \frac{5}{x} + \frac{1}{x^2}\right)
\]

\[
= \frac{\lim_{x \to \infty} 1 + \lim_{x \to \infty} \frac{3}{x} + \lim_{x \to \infty} \frac{4}{x^2}}{\lim_{x \to \infty} 2 - \lim_{x \to \infty} \frac{5}{x} + \lim_{x \to \infty} \frac{1}{x^2}}
\]
We conclude that this function has a horizontal asymptote at \( y = \frac{1}{2} \).

Many functions do not have horizontal asymptotes. For example, the function \( f(x) = x \) does not approach any finite value as \( x \) becomes infinite; that is, \( f(x) \) itself is becoming infinite as well. We can use the notation of limits to describe this behavior as well.

**Definition 1.15 (Infinite Limits at Infinity)** Let \( f(x) \) be a function that is defined on an interval of the form \((a, \infty)\) for some number \( a \). We say that the limit of \( f(x) \) as \( x \) approaches \( \infty \) is equal to \( \infty \), and write

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

if for any number \( N \), there exists a number \( M \) such that \( f(x) > N \) whenever \( x > M \). We say that the limit of \( f(x) \) as \( x \) approaches \( -\infty \) is equal to \( -\infty \), and write

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

if for any number \( N \), there exists a number \( M \) such that \( f(x) < N \) whenever \( x > M \).

Similarly, if \( f(x) \) is defined on an interval of the form \((-\infty, a)\) for some number \( a \), then we say that the limit of \( f(x) \) as \( x \) approaches \(-\infty \) is equal to \( \infty \), and write

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

if for any number \( N \), there exists a number \( M \) such that \( f(x) > N \) whenever \( x < M \). We say that the limit of \( f(x) \) as \( x \) approaches \(-\infty \) is equal to \( -\infty \), and write

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

if for any number \( N \), there exists a number \( M \) such that \( f(x) < N \) whenever \( x < M \).
Example 1.60 For \( f(x) = x \), we have the function value \( f(x) \) being equal to the corresponding \( x \)-value, and therefore

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

On the other hand, if \( f(x) = x^2 \), then \( f(x) \) approaches \( \infty \) as \( x \) approaches \( \infty \) or \( -\infty \); that is,

\[
\lim_{x \to \infty} x^2 = \infty, \quad \lim_{x \to -\infty} x^2 = \infty.
\]

Finally, we can determine from a graph or by substituting \( x \)-values that

\[
\lim_{x \to \infty} x^3 = \infty, \quad \lim_{x \to -\infty} x^3 = -\infty.
\]

It is important to keep in mind that the Limit Laws cannot be applied to infinite limits at infinity, since arithmetic operations are not defined for \( \infty \) or \( -\infty \).

Example 1.61 Compute

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

Solution Both the numerator and denominator become infinite as \( x \) becomes infinite, so we divide both the numerator and denominator by the highest power of \( x \) in the entire fraction, which is \( x^2 \). We can then use the Limit Laws to compute the limit. We have

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

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

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

\[
= \lim_{x \to \infty} \left( 2 + \frac{3}{x} + \frac{1}{x^2} \right)
\]

\[
= \lim_{x \to \infty} \left( 3 + \frac{4}{x} - \frac{5}{x^2} \right)
\]

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

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

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

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

\[
= 2 + 3(0) + 0
\]

\[
= 3 + 4(0) - 5(0)
\]

\[
= \frac{2}{3}.
\]
Note that we have used the fact that
\[ \lim_{x \to \infty} \frac{1}{x^r} = 0 \quad (1.170) \]
for any positive rational number \( r \).

**Example 1.62** Compute
\[ \lim_{x \to \infty} \sqrt{x^2 + x + 1} - x. \quad (1.171) \]

**Solution** As \( x \) approaches \( \infty \), each term individually becomes infinite, but \( \infty - \infty \) is undefined. Therefore, we instead multiply and divide by the conjugate \( \sqrt{x^2 + x + 1} + x \). Then, we obtain a quotient and can apply the Limit Laws. We have
\[
\lim_{x \to \infty} \sqrt{x^2 + x + 1} - x = \lim_{x \to \infty} \frac{(\sqrt{x^2 + x + 1} - x)(\sqrt{x^2 + x + 1} + x)}{\sqrt{x^2 + x + 1} + x} \\
= \lim_{x \to \infty} \frac{(\sqrt{x^2 + x + 1} - x)^2 - x^2}{\sqrt{x^2 + x + 1} + x} \\
= \lim_{x \to \infty} \frac{x^2 + x + 1 - x^2}{\sqrt{x^2 + x + 1} + x} \\
= \lim_{x \to \infty} \frac{x + 1}{\sqrt{x^2 + x + 1} + x} \\
= \lim_{x \to \infty} \frac{1 + \frac{1}{x}}{\sqrt{\frac{x^2 + x + 1}{x^2}} + 1} \\
= \lim_{x \to \infty} \frac{1 + \frac{1}{x}}{\sqrt{1 + \frac{1}{x} + \frac{1}{x^2}} + 1} \\
= \lim_{x \to \infty} \frac{1 + \frac{1}{x}}{\sqrt{1 + \frac{1}{x} + \frac{1}{x^2}} + 1}
\]
\[1.7. \text{LIMITS INVOLVING INFINITY} \]

\[
\begin{align*}
&= \frac{\lim_{x \to \infty} (1 + \frac{1}{x})}{\lim_{x \to \infty} \left( \sqrt{1 + \frac{1}{x} + \frac{1}{x^2} + 1} \right)} \\
&= \frac{\lim_{x \to \infty} 1 + \lim_{x \to \infty} \frac{1}{x}}{\lim_{x \to \infty} \sqrt{1 + \frac{1}{x} + \frac{1}{x^2} + \lim_{x \to \infty} 1}} \\
&= \frac{\lim_{x \to \infty} 1 + \lim_{x \to \infty} \frac{1}{x}}{\sqrt{\lim_{x \to \infty} 1 + \lim_{x \to \infty} \frac{1}{x} + \lim_{x \to \infty} \frac{1}{x^2} + \lim_{x \to \infty} 1}} \\
&= \frac{1 + 0}{\sqrt{1 + 0 + 0 + 1}} \\
&= \frac{1}{\sqrt{1 + 1}} \\
&= \frac{1}{1 + 1} \\
&= \frac{1}{2}.
\end{align*}
\]

Note that one must be careful in using the Limit Laws regarding exponents. In this case, we could use

\[
\lim_{x \to \infty} \sqrt{f(x)} = \sqrt{\lim_{x \to \infty} f(x)} \quad (1.172)
\]

because \(\lim_{x \to \infty} f(x)\) is finite. However, we cannot use this Limit Law in any situation that would involve exponentiating \(\infty\) or \(-\infty\). \(\blacksquare\)
2.1 Derivatives and Rates of Change

In the previous chapter, we learned that the instantaneous rate of change of a function $f(x)$ at a point $x = a$ was equal to the slope of a line that passed through the point $(a, f(a))$ and was tangent to the curve $y = f(x)$. In this section we will use the notation of limits to more precisely define a tangent line and the corresponding instantaneous rate of change.

2.1.1 Velocities

Consider a function $f(t)$, where the independent variable $t$ denotes time and $f(t)$ denotes the position of an object at time $t$. From previous discussion, we learned that we could compute the instantaneous velocity of the object at a particular time $t_0$ by computing its average velocity over an interval $(t_0, t_0 + h)$ for some number $h$, using the simple formula

$$v = \frac{d}{t} = \frac{f(t_0 + h) - f(t_0)}{(t_0 + h) - t_0} = \frac{f(t_0 + h) - f(t_0)}{h}$$

(2.1)

where $v$ is the average velocity from time $t_0$ to time $t_0 + h$, $t$ is the elapsed time, and $d$ is the distance traveled during this interval of time. As we chose smaller and smaller values of $h$, the interval $(t_0, t_0 + h)$ shrunk to the single point $t_0$, and the average velocity over the interval converged to a particular value, which we then defined to be the instantaneous velocity at time $t_0$. We now use the notation of limits to more precisely define this concept.

**Definition 2.1** Let the function $f(t)$ denote the position of an object at time $t$. Then the **instantaneous velocity** of the object at time $t_0$, denoted
by $v(t_0)$, is
\[
v(t_0) = \lim_{h \to 0} \frac{f(t_0 + h) - f(t_0)}{h},
\] (2.2)
provided that this limit exists.

**Example 2.1** Suppose that the position of an object at time $t$, where $t$ is in seconds and position is measured in feet, is described by the function $f(t) = t^2$. We wish to compute its velocity at time $t = 2$, which we denote by $v(2)$. Using the definition of instantaneous velocity, along with the limit laws introduced in Section 2.2, we obtain

\[
v(2) = \lim_{h \to 0} \frac{f(2 + h) - f(2)}{h}
= \lim_{h \to 0} \frac{(2 + h)^2 - 2^2}{h}
= \lim_{h \to 0} \frac{4 + 4h + h^2 - 4}{h}
= \lim_{h \to 0} \frac{4h + h^2}{h}
= \lim_{h \to 0} 4 + h
= \lim_{h \to 0} 4 + \lim_{h \to 0} h
= 4.
\]

We conclude that the velocity at time $t = 2$ is 4 ft/s. $\Box$

### 2.1.2 Tangents

In Section 1.3 we also discussed a geometric interpretation of the concepts of average velocity and instantaneous velocity. Recall that the average velocity over the interval $(t_0, t_0 + h)$ is equal to the slope of the secant line that passes through the points $(t_0, f(t_0))$ and $(t_0 + h, f(t_0 + h))$. As $h$ approaches 0, the corresponding secant line converges to a line that is tangent to the curve $y = f(t)$ at the point $(t_0, f(t_0))$; that is, it touches the graph of $f$ at this point but does not cross the graph. The slope of this tangent line is equal to the instantaneous velocity at time $t_0$.

We can now give a precise definition of a tangent line using the notation of limits. In the following definition, we use the letter $x$ for the independent variable instead of $t$, and the letter $a$ to denote the point of tangency instead of the symbol $t_0$. This notation is used to emphasize the fact that the given
2.1. DERIVATIVES AND RATES OF CHANGE

function $f$ need not be a function of time; the independent variable can have any interpretation.

**Definition 2.2** The **tangent line** to the curve $y = f(x)$ at the point $(a, f(a))$ is the line with slope

$$m = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} \quad (2.3)$$

that passes through the point $(a, f(a))$, provided that the above limit $m$ exists.

If we let $x = a + h$, then $x$ approaches $a$ as $h$ approaches 0. We use this notation to give an alternative definition that is consistent with the definition given in the text.

**Definition 2.3** The **tangent line** to the curve $y = f(x)$ at the point $(a, f(a))$ is the line with slope

$$m = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \quad (2.4)$$

that passes through the point $(a, f(a))$, provided that the above limit $m$ exists.

**Example 2.2** We again consider the function $f(t) = t^2$ and compute the equation of the tangent line at the point $(2, 4)$. To compute the slope, denoted by $m$, we need to evaluate

$$m = \lim_{t \to 2} \frac{f(t) - f(2)}{t - 2}. \quad (2.5)$$

Using the limit laws, we obtain

$$m = \lim_{t \to 2} \frac{f(t) - f(2)}{t - 2}
= \lim_{t \to 2} \frac{t^2 - 2^2}{t - 2}
= \lim_{t \to 2} \frac{t^2 - 4}{t - 2}
= \lim_{t \to 2} \frac{(t + 2)(t - 2)}{t - 2}
= \lim_{t \to 2} (t + 2)
= 2 + 2
= 4.$$
Therefore, the tangent line is the line with slope 4 that passes through the point \((2, 4)\). It follows that the equation of the tangent line, in point-slope form, is

\[ y - 4 = 4(t - 2) \]  

or, in slope-intercept form,

\[ y = 4t - 4. \]  

The tangent line is illustrated in Figure 2.1.

Figure 2.1: Tangent line of \(y = t^2\) at the point \((2, 4)\). The point of tangency is indicated by the red circle.

2.1.3 Other Rates of Change

We have seen that the slope of the tangent line of a function \(y = f(t)\) at a point \(t_0\) is equal to the instantaneous velocity at time \(t_0\) of the object whose position at time \(t\) is given by \(f(t)\). However, since velocity at a particular time is defined to be the instantaneous rate of change of position with respect
to time, and since the definition of the tangent line is independent of any interpretation of the independent variable or dependent variable, it is natural to equate the slope of the tangent line of $y = f(x)$ at $(a, f(a))$ with the instantaneous rate of change of $y$ with respect to $x$ at $x = a$, regardless of the interpretation of $x$ and $y$. This leads to the following definitions.

**Definition 2.4** Let $y = f(x)$. The **increment** of $x$ as $x$ changes from $a$ to $b$, denoted by $\Delta x$, is given by

$$\Delta x = b - a. \quad (2.8)$$

Similarly, the **increment** of $y$ as $x$ changes from $a$ to $b$, denoted by $\Delta y$, is given by

$$\Delta y = f(b) - f(a). \quad (2.9)$$

The **average rate of change of $y$ with respect to $x$ over the interval** $[a, b]$ is given by the **difference quotient**

$$\frac{\Delta y}{\Delta x} = \frac{f(b) - f(a)}{b - a}. \quad (2.10)$$

The **instantaneous rate of change of $y$ with respect to $x$ at $x = a$ is given by**

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{b \to a} \frac{f(b) - f(a)}{b - a}. \quad (2.11)$$

**Example 2.3** Suppose that one is studying the temperature within a rod that is 4 m long. The temperature $T$ can be described by a function $u(x)$, where $x$ denotes the position on the rod at which the temperature is measured. In this example, we assume that the temperature is held fixed at 0°C at both ends of the rod, and that the temperature the point $x$ is given by

$$u(x) = 1 - \frac{(x - 2)^2}{4}. \quad (2.12)$$

Note that for this choice of $u$, $u(0) = u(4) = 0$, as intended. We wish to compute the instantaneous rate of change at the point that is 1 m from the right endpoint the rod; i.e., at $x = 3$. From the definition of the instantaneous rate of change of $u$ with respect to $x$ at $x = 3$, we obtain

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{x \to 3} \frac{u(x) - u(3)}{x - 3}$$

$$= \lim_{x \to 3} \frac{1 - \frac{(x-2)^2}{4} - \left(1 - \frac{(3-2)^2}{4}\right)}{x - 3}$$
\[
\lim_{x \to 3} \frac{1 - \frac{(x-2)^2}{4} - 1 + \frac{(3-2)^2}{4}}{x - 3} \\
\lim_{x \to 3} \frac{(x-2)^2 + (3-2)^2}{4} \\
\frac{1}{4} \lim_{x \to 3} \frac{-(x-2)^2 + (3-2)^2}{x - 3} \\
\frac{1}{4} \lim_{x \to 3} \frac{-(x^2 - 4x + 4) + 1}{x - 3} \\
\frac{1}{4} \lim_{x \to 3} \frac{-(x^2 - 4x + 4) - 1}{x - 3} \\
\frac{1}{4} \lim_{x \to 3} \frac{x^2 - 4x + 3}{x - 3} \\
\frac{1}{4} \lim_{x \to 3} \frac{(x - 1)(x - 3)}{x - 3} \\
\frac{1}{4} \lim_{x \to 3} x - 1 \\
\frac{1}{4} \cdot 2 \\
\frac{1}{2}.
\]

We conclude that at \( x = 3 \), the temperature is decreasing at a rate of one-half degree Celsius per meter as we traverse the rod from left to right, since \( x \) denotes position along the rod and increases from left to right. □

As in the case where the dependent variable denotes position and the independent variable denotes time, the average rate of change of \( y \) with respect to \( x \) over \([a, b]\) is equal to the slope of the secant line passing through the points \((a, f(a))\) and \((b, f(a))\). Likewise, the instantaneous rate of change of \( y \) with respect to \( x \) at \( x = a \) is equal to the slope of the tangent line to the curve \( y = f(x) \) at the point \((a, f(a))\).

Now that we have a precise definition for the instantaneous rate of change of a quantity \( y \) with respect to another quantity \( x \), where \( x \) and \( y \) are related by the equation \( y = f(x) \) for some function \( f \), we can turn our attention to the actual computation of this rate of change for a given function \( f \). This will be the subject of the next few sections.
2.1. DERIVATIVES AND RATES OF CHANGE

2.1.4 The Derivative

In the previous section we gave a precise definition of the instantaneous rate of change of a quantity \( y \) with respect to another quantity \( x \), at a particular value of \( x \). When the quantities \( x \) and \( y \) are related by an equation of the form \( y = f(x) \), it is certainly convenient to describe this rate of change in terms of the function \( f \), and because the instantaneous rate of change is so commonplace, it is practical to assign a concise name and notation to it, which we do now.

**Definition 2.5 (Derivative)** The **derivative** of a function \( f(x) \) at \( x = a \), denoted by \( f'(a) \), is

\[
f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}, \tag{2.13}
\]

provided that the above limit exists. When this limit exists, we say that \( f \) is **differentiable** at \( a \).

As with the definition of a tangent line in Section 2.1, we can set \( x = a + h \) and obtain the following alternative, but equivalent, definition of the derivative.

**Definition 2.6 (Derivative, alternative definition)** The **derivative** of a function \( f(x) \) at \( x = a \), denoted by \( f'(a) \), is

\[
f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}, \tag{2.14}
\]

provided that the above limit exists.

**Remark** Given a function \( f(x) \) that is differentiable at \( x = a \), the following numbers are all equal:

- the derivative of \( f \) at \( x = a \), \( f'(a) \),
- the slope of the tangent line of \( f \) at the point \((a, f(a))\), and
- the instantaneous rate of change of \( y = f(x) \) with respect to \( x \) at \( x = a \).

This can be seen from the fact that all three numbers are defined in the same way.
Example 2.4  A ball thrown upward at a velocity of 40 ft/s has an height of \( f(t) = 40t - 16t^2 \) feet after \( t \) seconds. We can verify that its initial velocity is 40 ft/s by computing the instantaneous rate of change of \( y = f(t) \) with respect to \( t \) at the time \( t = 0 \). This is equivalent to computing the derivative of \( f(t) \) at \( t = 0 \), which can be accomplished using the definition:

\[
\begin{align*}
  f'(0) &= \lim_{t \to 0} \frac{f(t) - f(0)}{t - 0} \\
  &= \lim_{t \to 0} \frac{40t - 16t^2 - 0}{t} \\
  &= \lim_{t \to 0} 40 - 16t \\
  &= 40.
\end{align*}
\]

We also note that the slope of the tangent line of \( f(t) \) at the point \((0, 0)\) is equal to 40. □

2.2 The Derivative as a Function

Because the derivative of a function \( f(x) \) at \( x = a \) is defined to be the instantaneous rate of change of \( y = f(x) \) with respect to \( x \) at \( x = a \), it follows that the derivative itself is a function that associates the number \( f'(x) \) with \( x \) for every \( x \) in the domain of \( f \) at which \( f'(x) \) is defined. We now make precise this notion of the derivative as a function.

Definition 2.7 (Derivative as a function) Given a function \( f \), the derivative of \( f \), denoted by \( f' \), is a function defined at every number \( x \) such that \( x \) is in the domain of \( f \), and \( f \) is differentiable at \( x \). The value of \( f' \) at \( x \) is

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

Example 2.5 Consider \( f(x) = x^2 \). By the definition of the derivative,

\[
\begin{align*}
  f'(x) &= \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \\
  &= \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h} \\
  &= \lim_{h \to 0} \frac{x^2 + 2hx + h^2 - x^2}{h} \\
  &= \lim_{h \to 0} \frac{2hx + h^2}{h} \\
  &= 2x.
\end{align*}
\]
Therefore, we say that \( f'(x) = 2x \) is the derivative of \( f(x) = x^2 \).

Example 2.6 Given \( f(x) = x^3 \), compute the derivative of \( f(x) \) at \( x = 2 \), \( x = 4 \), and \( x = 5 \).

Solution We examine two approaches. The first uses the notion of the derivative as a number. Recall that we defined the derivative of \( f(x) \) at \( x = a \), denoted \( f'(a) \), by

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

(2.16)

provided this limit exists. Using this definition, we have

\[
f'(2) = \lim_{h \to 0} \frac{f(2 + h) - f(2)}{h} = \lim_{h \to 0} \frac{(2 + h)^3 - 2^3}{h} = \lim_{h \to 0} \frac{(8 + 12h + 6h^2 + h^3) - 8}{h} = \lim_{h \to 0} 12 + 6h + h^2 = 12.
\]

We can repeat this process for \( x = 4 \) and \( x = 5 \), but it can become tedious to compute a separate limit for each \( x \)-value at which we want the value of the derivative. It is more practical to use the notion of the derivative of \( f(x) \) as a function \( f'(x) \) whose value at \( x \) is given by

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

(2.17)

provided the limit exists. This leads to the alternative approach of obtaining the function \( f'(x) \) and simply evaluating this function at \( x = 2 \), \( x = 4 \), and \( x = 5 \). Using the above definition of the function \( f'(x) \), we obtain

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

It follows that

\[ f'(2) = 3(2^2) = 12, \quad f'(4) = 3(4^2) = 48, \quad f'(5) = 3(5^2) = 75. \quad (2.18) \]

Using this approach to obtain the derivative at several points, we only needed to compute one limit, that represented a function rather than a number, and then evaluate that function at each point, instead of computing a separate limit for each point. □

In upcoming sections we will develop several rules that can be used to easily compute the function \( f'(x) \) that is the derivative of a given function \( f(x) \). As a result, it will rarely be necessary to use the above definition directly. As is the case with the definition of a limit and the definition of continuity, the most practical use of the definition of the derivative is to compute the derivatives of abstract, general functions (such as an arbitrary polynomial, or a quotient of arbitrary functions) so that derivatives of specific functions can be computed much more easily. However, it is still necessary to know such definitions, since there are difficult problems to which they are directly applicable.

### 2.2.1 Notation

There are a number of ways to denote the derivative of a function \( f \). While some notations are more common than others, it is useful to be aware of as many of them as possible in order to more easily comprehend any writings in which the derivative is used. The following are all equivalent:

- \( f'(x) \)
- \( df/dx \)
2.2. THE DERIVATIVE AS A FUNCTION

- \( \frac{d}{dx} f(x) \). The expression \( \frac{d}{dx} \) is called a differentiation operator. This notation is typically used to emphasize the fact that \( \frac{d}{dx} f(x) \) is a function that is the result of an operation, called differentiation, that computes the derivative. Other notations for the differentiation operator are \( D \) and \( D_x \), though they will not be used in this book.

- \( y' \), where \( y = f(x) \).

- \( \frac{dy}{dx} \), where \( y = f(x) \). To denote the value of the derivative at a particular point \( x = a \), one uses the notation

\[
\left. \frac{dy}{dx} \right|_{x=a}.
\]  

(2.19)

2.2.2 Differentiability and Continuity

Consider a tangent line of a function \( f \) at a point \( (a, f(a)) \). When we consider that this tangent line is the limit of secant lines that can cross the graph of \( f \) at points on either side of \( a \), it seems impossible that \( f \) can fail to be continuous at \( a \). The following result confirms this: a function that is differentiable at a given point (and therefore has a tangent line at that point) must also be continuous at that point.

**Theorem 2.1** If \( f \) is differentiable at \( a \), then \( f \) is continuous at \( a \).

**Proof** Since \( f \) is differentiable at \( a \), it follows that the limit

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

(2.20)

exists. In order to prove that \( f \) is continuous at \( a \), we must show that

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

(2.21)

Using the limit laws, we proceed as follows:

\[
\lim_{x \to a} f(x) = \lim_{x \to a} [f(a) + (f(x) - f(a))] \\
= \lim_{x \to a} f(a) + \lim_{x \to a} [f(x) - f(a)] \\
= f(a) + \lim_{x \to a} \frac{f(x) - f(a)}{x - a} (x - a) \\
= f(a) + \left[ \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \right] \cdot \left[ \lim_{x \to a} (x - a) \right] \\
= f(a) + f'(a) \cdot 0 \\
= f(a)
\]
and the proof is complete. □

It is important to keep in mind, however, that the converse of the above statement, “if \( f \) is continuous, then \( f \) is differentiable”, is not true. It is actually very easy to find examples of functions that are continuous at a point, but fail to be differentiable at that point.

**Example 2.7** The best-known example of a function that is continuous and not differentiable at a point is the absolute value function \( f(x) = |x| \). This function is continuous for all \( x \), but is not differentiable at \( x = 0 \). To see this, we compute the left-hand and right-hand limits

\[
\lim_{x \to 0^-} \frac{f(x) - f(0)}{x - 0}, \quad \lim_{x \to 0^+} \frac{f(x) - f(0)}{x - 0}. \tag{2.22}
\]

For the left-hand limit, we consider \( x < 0 \), and therefore

\[
\lim_{x \to 0^-} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0^-} \frac{|x| - |0|}{x} = \lim_{x \to 0^-} \frac{-x}{x} = -1, \tag{2.23}
\]

whereas for the right-hand limit, we consider \( x > 0 \), and therefore

\[
\lim_{x \to 0^+} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0^+} \frac{|x| - |0|}{x} = \lim_{x \to 0^+} \frac{x}{x} = 1. \tag{2.24}
\]

Since the left-hand and right-hand limits do not agree, we conclude that

\[
\lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} \tag{2.25}
\]
does not exist, and therefore \( |x| \) is not differentiable at \( x = 0 \). □

In general, a function can fail to be differentiable at a point for one of the following reasons:

- If the graph of \( f \) has a sharp corner at a point, as \( |x| \) does, then, as seen in the above example, the left-hand and right-hand limits of the slopes of secant lines do not agree, so the function has no tangent line at that point and is therefore not differentiable there.

- If a function is not continuous at a point, it cannot be differentiable at that point. This is because if it were differentiable, then it must be continuous, which is a contradiction.
2.3. **BASIC DIFFERENTIATION FORMULAS**

- If a function has a vertical tangent line at a point, due to the graph becoming a vertical line or the function having an infinite limit, then the function is not differentiable at that point because the derivative must have a finite value.

As an extreme example, it is known that there is a function that is continuous everywhere, but is not differentiable *nowhere*.

**Example 2.8** The function

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{2^n} \cos(12^{-n} \pi x), \quad (2.26)$$

constructed by Karl Weierstrass in 1872, is continuous at every $x$, but is not differentiable at any $x$. □

### 2.3 Basic Differentiation Formulas

In the previous section, we defined the derivative of a function $f(x)$, denoted by $f'(x)$ or $df(x)/dx$, to be the function whose value at any point $a$ is equal to the instantaneous rate of change of $y = f(x)$ with respect to $x$ at $x = a$. In other words,

$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}, \quad (2.27)$$

or, alternatively,

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}. \quad (2.28)$$

While this definition can always be used to differentiate (i.e., compute the derivative of) a given function, this is not practical, just as it is not practical to use the definition of a limit to compute the limit of a given function. Instead, we will use the definition to establish some differentiation rules that can be used to more easily differentiate a wide variety of functions.

We begin by computing the derivatives of some very simple functions.

First, we consider the constant function $f(x) = c$. Applying the definition, we obtain

$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} = \lim_{h \to 0} \frac{c - c}{h} = \lim_{h \to 0} \frac{0}{h} = \lim_{h \to 0} 0 = 0. \quad (2.29)$$

Intuitively, this makes sense because if $y = f(x)$, where $f(x)$ is equal to a constant $c$ for all $x$, then the value of $y$ does not depend on $x$ at all, and so
its instantaneous rate of change with respect to $x$ should always be zero. In summary,

$$\frac{d}{dx}(c) = 0. \quad (2.30)$$

Next, we turn to the identity function $f(x) = x$. Applying the definition to compute $f'(a)$, we obtain

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

(2.31)

Intuitively, if $y = x$, any change in $x$ implies the same change in $y$, so the instantaneous rate of change of $y$ with respect to $x$ should always be equal to 1. In summary,

$$\frac{d}{dx}(x) = 1. \quad (2.32)$$

Now, we will use the definition to establish some rules that can be used to obtain the derivatives of complicated functions in terms of the derivatives of simpler functions. These rules are direct consequences of the limit laws introduced in Section 1.5. First, we consider the derivative of the sum of two functions, $f(x) + g(x)$. If we already know the derivatives of $f$ and $g$, what is the derivative of $f + g$? Applying the definition and the limit law for sums, we obtain

$$\frac{d}{dx}[f(x) + g(x)] = \lim_{h \to 0} \frac{[f(x+h) + g(x+h)] - [f(x) + g(x)]}{h} = \lim_{h \to 0} \frac{[f(x+h) - f(x)] + [g(x+h) - g(x)]}{h} = \lim_{h \to 0} \left[ \frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h} \right] = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = \frac{d}{dx}f(x) + \frac{d}{dx}g(x).$$

In summary, we have obtained the following result.

**Theorem 2.2** (Sum Rule) If $f$ and $g$ are differentiable at $x$, then $f + g$ is differentiable at $x$ and

$$\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x). \quad (2.33)$$
2.3. BASIC DIFFERENTIATION FORMULAS

The limit law for the difference of functions can be applied in the same manner to obtain the following rule.

**Theorem 2.3** (Difference Rule) If \( f \) and \( g \) are differentiable at \( x \), then \( f - g \) is differentiable at \( x \) and

\[
\frac{d}{dx} [f(x) - g(x)] = \frac{d}{dx} f(x) - \frac{d}{dx} g(x).
\] (2.34)

**Example 2.9** Let \( f(x) = 2x \). Then, writing \( f(x) = x + x \) and applying the Sum Rule, we obtain

\[
f'(x) = \frac{d}{dx} x + \frac{d}{dx} x = 1 + 1 = 2.
\] (2.35)

\( \square \)

**Example 2.10** Let \( f(x) = 3x \). Then, writing \( f(x) = 2x + x \) and applying the Sum Rule, we obtain

\[
f'(x) = \frac{d}{dx} [2x] + \frac{d}{dx} x = 2 + 1 = 3.
\] (2.36)

\( \square \)

From the previous two examples, we have seen that whereas the derivative of \( x \) is equal to 1, the derivative of \( 2x \) is 2, while the derivative of \( 3x \) is 3. It is natural to ask whether scaling a function by a constant \( c \) scales its derivative by the same constant \( c \), as these examples suggest. We can use the definition to answer this question in the general case. We have

\[
\frac{d}{dx} [cf(x)] = \lim_{h \to 0} \frac{cf(x + h) - cf(x)}{h}
= \lim_{h \to 0} c \frac{f(x + h) - f(x)}{h}
= \left[ \lim_{h \to 0} c \right] \left[ \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} \right]
= c \frac{d}{dx} f(x).
\]

We have just proven the following theorem.

**Theorem 2.4** (Constant Multiple Rule) If \( f \) is differentiable at \( x \) and \( c \) is a constant, then \( cf \) is differentiable at \( x \) and

\[
\frac{d}{dx} [cf(x)] = c \frac{d}{dx} f(x).
\] (2.37)
Example 2.11 Let \( f(x) = 3x + 2 \). Applying the Sum Rule and the Constant Multiple Rule, we obtain
\[
f'(x) = \frac{d}{dx}[3x + 2] = \frac{d}{dx}[3x] + \frac{d}{dx}2 = 3 \frac{d}{dx}x + 0 = 3 \cdot 1 = 3. \tag{2.38}
\]
\Box

Example 2.12 Let \( f(x) = mx + b \), where \( m \) and \( b \) are constants. Then \( f \) is a linear function, whose graph is a line with slope \( m \) and \( y \)-intercept \( b \). Using the Sum Rule and the Constant Multiple Rule, we obtain
\[
f'(x) = \frac{d}{dx}[mx + b] = \frac{d}{dx}[mx] + \frac{d}{dx}b = m \frac{d}{dx}x + 0 = m \cdot 1 = m. \tag{2.39}
\]
In other words, the derivative of a linear function is always equal to its slope. This makes sense, because if \( y = mx + b \), then the slope \( m \) is equal to the rate of change of \( y \) with respect to \( x \): every unit of “run” corresponds to \( m \) units of “rise”. \Box

2.3.1 Derivatives of Trigonometric Functions

Using the differentiation rules introduced in the previous section, we can now differentiate any algebraic function. We now turn our attention to the other category of continuous functions used in this book, trigonometric functions, so that we can efficiently differentiate those functions as well.

We begin with the function \( f(x) = \sin x \). Applying the definition of the derivative, we have
\[
f'(x) = \lim_{h \to 0} \frac{\sin(x + h) - \sin x}{h}. \tag{2.40}
\]
We use the trigonometric identity
\[
\sin(a + b) = \sin a \cos b + \cos a \sin b \tag{2.41}
\]
and use the limit laws to obtain
\[
f'(x) = \lim_{h \to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} = \lim_{h \to 0} \left[ \sin x \frac{\cos h - 1}{h} + \cos x \frac{\sin h}{h} \right] = \sin x \lim_{h \to 0} \frac{\cos h - 1}{h} + \cos x \lim_{h \to 0} \frac{\sin h}{h}.
\]
We now focus on the limit
\[
\lim_{h \to 0} \frac{\sin h}{h}.
\] (2.42)

Using right triangle trigonometry, it can be shown that for \(0 < h < \pi/2\),
\[
\sin h < h \quad \text{and} \quad h < \tan h.
\] (2.43)

These relations can be rearranged algebraically to obtain
\[
\cos h < \frac{\sin h}{h} < 1.
\] (2.44)

Since \(\lim_{h \to 0} \cos h = 1\), we can apply the Squeeze Theorem to conclude that
\[
\lim_{h \to 0^+} \frac{\sin h}{h} = 1.
\] (2.45)

However, the function \(g(h) = \frac{\sin(h)}{h}\) is an even function; that is, \(g(-h) = g(h)\). It follows that
\[
\lim_{h \to 0^-} g(h) = \lim_{h \to 0^+} g(h) = \lim_{h \to 0} g(h) = 1.
\] (2.46)

Therefore,
\[
\lim_{h \to 0} \frac{\sin h}{h} = 1.
\] (2.47)

Now, we examine the limit
\[
\lim_{h \to 0} \frac{\cos h - 1}{h}.
\] (2.48)

Using the half-angle formula
\[
\sin^2 x = \frac{1 - \cos 2x}{2}
\] (2.49)
yields
\[
\lim_{h \to 0} \frac{\cos h - 1}{h} = \lim_{h \to 0} -\frac{2\sin^2(h/2)}{h} \\
= -2 \lim_{h \to 0} \frac{\sin(h/2)\sin(h/2)}{h} \\
= -2 \lim_{h \to 0} \frac{\sin(h/2)/2}{h/2}
\]
Using these two limits, we finally obtain
\[
f'(x) = \frac{d}{dx}[\sin x] = \sin x \lim_{h \to 0} \frac{\cos h - 1}{h} + \cos x \lim_{h \to 0} \frac{\sin h}{h} = \sin x \cdot 0 + \cos x \cdot 1 = \cos x.
\]
In summary, we have the new differentiation rule,
\[
\frac{d}{dx} \sin x = \cos x.
\]
(2.50)

We can use a similar approach to compute the derivative of \( f(x) = \cos x \).
Using the definition of the derivative and the identity
\[
\cos(a + b) = \cos a \cos b - \sin a \sin b,
\]
we obtain
\[
f'(x) = \frac{d}{dx} [\cos x] = \lim_{h \to 0} \frac{\cos(x + h) - \cos x}{h} = \lim_{h \to 0} \frac{\cos x \cos h - \sin x \sin h - \cos x}{h} = \lim_{h \to 0} \left[ \cos x \frac{\cos h - 1}{h} - \sin x \frac{\sin h}{h} \right] = \cos x \lim_{h \to 0} \frac{\cos h - 1}{h} - \sin x \lim_{h \to 0} \frac{\sin h}{h} = \cos x \cdot 0 - \sin x \cdot 1 = -\sin x.
\]
In summary, we have the differentiation rule
\[
\frac{d}{dx} [\cos x] = -\sin x.
\]
(2.53)
2.4 The Product and Quotient Rules

The differentiation rules we have presented so far only allow us to compute the derivative of a linear function, as well as \( \sin x \) and \( \cos x \). We can expand our repertoire by establishing more rules. Since we know how to compute the derivative of a sum or difference of two functions, it is natural to consider the product of two functions as well. Using the definition, we proceed as follows:

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

\[
= \left[ \lim_{h \to 0} g(x+h) \right] \cdot \left[ \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \right] + \left[ \lim_{h \to 0} f(x) \right] \cdot \left[ \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \right] = g(x) \frac{d}{dx} [f(x)] + f(x) \frac{d}{dx} [g(x)].
\]

This yields the following rule.

**Theorem 2.5 (Product Rule)** If \( f \) and \( g \) are differentiable at \( x \), then \( fg \) is differentiable at \( x \) and

\[
\frac{d}{dx} [f(x)g(x)] = g(x) \frac{d}{dx} [f(x)] + f(x) \frac{d}{dx} [g(x)]. \tag{2.55}
\]

**Example 2.13** Let \( f(x) = (3x+2)(x+5) \). Then, using the Sum Rule and the Constant Multiple Rule, we have

\[
\frac{d}{dx} (3x+2) = 3, \quad \frac{d}{dx} (x+5) = 1. \tag{2.56}
\]

We can now use the Product Rule to obtain

\[
\frac{d}{dx} [f(x)] = (x+5) \frac{d}{dx} (3x+2) + (3x+2) \frac{d}{dx} (x+5)
\]
The derivative of $x^n$, where $n$ is a positive integer, is equal to $nx^{n-1}$. In fact, we have not only seen that this is the case for $n = 2, 3$ and 4, but also for $n = 0$ and $n = 1$, since the derivative of $x^1 = x$ is equal to $1 \cdot x^0 = 1$, while the derivative of $x^0 = 1$ is equal to $0 \cdot x^{-1} = 0$. This pattern does in fact hold in general, as we formally state now.

**Theorem 2.6 (Power Rule)** For any real number $n$,

$$
\frac{d}{dx}(x^n) = nx^{n-1}. 
$$

In an optional section at the conclusion of these notes, we will prove that this rule holds for the case where $n$ is a nonnegative integer. The case where $n$ is a negative integer will be proven in the next section, after we learn how to compute the derivative of the reciprocal of a function. The case where $n$ is not an integer will be proven in the next chapter, as it relies on concepts from transcendental functions. For now, we will accept the rule without proof.
Example 2.15 Let \( f(x) = \sqrt{x} \). Then, writing \( \sqrt{x} = x^{1/2} \), we can compute \( f'(x) \) using the Power Rule and obtain
\[
f'(x) = \frac{d}{dx}[\sqrt{x}] = \frac{d}{dx}[x^{1/2}] = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}. \tag{2.61}
\]
\( \Box \)

Example 2.16 Let \( f(x) = (x + 2)^3 \). We can compute \( f'(x) \) by expanding \((x + 2)^3\) and then applying the Sum Rule, the Constant Multiple Rule, and the Power Rule to obtain
\[
f'(x) = \frac{d}{dx}[(x + 2)^3] \\
= \frac{d}{dx}[x^3 + 6x^2 + 12x + 8] \\
= \frac{d}{dx}[x^3] + 6 \frac{d}{dx}[x^2] + 12 \frac{d}{dx}[x] + \frac{d}{dx}8 \\
= 3x^2 + 6(2x) + 12(1) + 0 \\
= 3x^2 + 12x + 12.
\]

In Section 3.6 we will learn how to compute derivatives of functions like this one without having to expand powers of functions. \( \Box \)

Using the rules introduced so far, we are now able to compute the derivative of any polynomial function, as well as some algebraic functions. Later in this section, we will learn how to compute the derivative of any algebraic function, including any rational function.

### 2.4.1 Proof of the Power Rule (optional)

We now present a proof of the Power Rule
\[
\frac{d}{dx}[x^n] = nx^{n-1}, \tag{2.62}
\]
where \( n \) is a nonnegative integer. We will use a very powerful technique called mathematical induction that is commonly used not only in proofs but also in the development of computer algorithms.

We proceed as follows: we want to prove that this rule holds for all \( n \) where \( n \) is a nonnegative integer, so we first prove that it holds for the smallest nonnegative integer, \( n = 0 \). This case, in the context of induction, is called the basis step. Then, we prove that if the rule is true when \( n \) is equal
to any nonnegative integer $k$, then it must be true for $n = k + 1$. This step is called the induction step. The proof of the base case and the induction step will actually be sufficient to prove that the rule is true for all nonnegative integers. To see this, intuitively, think of an infinitely long sequence of dominos, where each domino corresponds to a nonnegative integer, and the domino corresponding to an integer $n$ is knocked down if the Power Rule is true for $n$. Proving the base case is equivalent to knocking over the first domino. In the induction case, it is shown that if any domino is knocked down, the next domino will be knocked down as well. Putting these two statements together, we can knock over all of the dominos.

We now proceed with the proof. For the base case of $n = 0$, we have

$$\frac{d}{dx}[x^0] = \frac{d}{dx}[1] = 0,$$

and

$$0 \cdot x^{0-1} = 0,$$

so we can conclude that

$$\frac{d}{dx}[x^0] = 0 \cdot x^{0-1},$$

and therefore the base case holds: the Power Rule is true for $n = 0$. Now, for the induction step, we assume that the Power Rule is true for $n = k$ and prove that it must be true for $n = k + 1$. We write $x^{k+1} = x^k x$ and use the Product Rule to obtain

$$\frac{d}{dx}[x^{k+1}] = \frac{d}{dx}[x^k x]$$

$$= x \frac{d}{dx}[x^k] + x^k \frac{d}{dx}x$$

$$= x(kx^{k-1}) + x^k \cdot 1$$

$$= kx^k + x^k$$

$$= (k + 1)x^k.$$

### 2.4.2 Differentiation of General Algebraic Functions

Earlier in this section, we learned a number of differentiation rules that enabled us to efficiently compute the derivative of any polynomial function, as well as some simple algebraic functions such as $f(x) = \sqrt{x}$. However, we have yet to learn how to compute the derivative of the quotient of two functions, or a function raised to a power, so we will address these problems now.
First, we consider the reciprocal of a function; that is, we wish to compute the derivative of \(1/g(x)\) for some function \(g\), where \(g\) is differentiable and nonzero at \(x\). Using the definition of the derivative, we obtain

\[
\frac{d}{dx} \left( \frac{1}{g(x)} \right) = \lim_{h \to 0} \frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} = \lim_{h \to 0} \frac{g(x) - g(x+h)}{h} \frac{1}{g(x+h)g(x)} \\
= \lim_{h \to 0} \frac{g(x) - g(x+h)}{h} \cdot \lim_{h \to 0} \frac{1}{g(x+h)g(x)} \\
= -\left[ \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \right] \cdot \left[ \lim_{h \to 0} \frac{1}{g(x+h)g(x)} \right] \\
= -\left[ \frac{d}{dx} g(x) \right] \cdot \left[ \frac{1}{[g(x)]^2} \right] \\
= -\frac{d}{dx} \left( \frac{1}{g(x)} \right) \cdot \left[ \frac{1}{[g(x)]^2} \right].
\]

**Example 2.17** Suppose that \(f(x) = 1/x\). Then, by the above discussion,

\[
f'(x) = \frac{d}{dx} \left( \frac{1}{x} \right) = -\frac{d}{dx} \frac{x}{x^2} = -\frac{1}{x^2}. \tag{2.66}
\]

Now, suppose that \(f(x) = x^{-n}\), where \(n\) is a nonnegative integer. Writing \(f(x) = 1/(x^n)\) and using the formula for the derivative of a reciprocal in conjunction with the Power Rule, we obtain

\[
f'(x) = \frac{d}{dx} \left( \frac{1}{x^n} \right) = -\frac{d}{dx} \left( \frac{x^n}{[x^n]^2} \right) = -\frac{nx^{n-1}}{x^{2n}} = -nx^{n-1-2} = -nx^{-n-1}. \tag{2.67}
\]

We have just proven that the Power Rule

\[
\frac{d}{dx} [x^n] = nx^{n-1} \tag{2.68}
\]

holds for negative values of \(n\). □

Now, we will compute the derivative of a general quotient \(f(x)/g(x)\), where \(g(x) \neq 0\). Applying the Product Rule, we obtain

\[
\frac{d}{dx} \left[ \frac{f(x)}{g(x)} \right] = \frac{d}{dx} \left[ f(x) \frac{1}{g(x)} \right]
\]
We have just proven the following result.

**Theorem 2.7 (Quotient Rule)** If \( f \) and \( g \) are differentiable at \( x \) and \( g(x) \neq 0 \), then \( f/g \) is differentiable at \( x \) and

\[
\frac{d}{dx} \left( \frac{f(x)}{g(x)} \right) = \frac{g(x) \frac{d}{dx} [f(x)] - f(x) \frac{d}{dx} [g(x)]}{[g(x)]^2}.
\]  (2.69)

**Example 2.18** We will compute the derivative of

\[ f(x) = \frac{x^3 - 1}{x^2 + 1}. \]  (2.70)

Using the Quotient Rule, we obtain

\[
f'(x) = \frac{d}{dx} \left[ \frac{x^3 - 1}{x^2 + 1} \right] = \frac{(x^2 + 1) \frac{d}{dx} [x^3 - 1] - (x^3 - 1) \frac{d}{dx} [x^2 + 1]}{(x^2 + 1)^2}
= \frac{(x^2 + 1)(3x^2) - (x^3 - 1)(2x)}{(x^2 + 1)^2}
= \frac{3x^4 + 3x^2 - 2x^4 - 2x}{(x^2 + 1)^2}
= \frac{x^4 + 3x^2 + 2x}{(x^2 + 1)^2}.
\]

\[ \square \]

Now we will learn how to compute the derivative of a function of the form \( g(x) = [f(x)]^n \), where \( n \) is a nonnegative integer. For \( n = 0 \), we have \( g(x) = 1 \), so \( g'(x) = 0 \). For \( n = 1 \), we have \( g(x) = f(x) \), so \( g'(x) = f'(x) \). For \( n = 2 \), we have \( g(x) = [f(x)]^2 \), so, by the Product Rule,

\[ g'(x) = \frac{d}{dx} \left[ [f(x)]^2 \right] \]
2.4. THE PRODUCT AND QUOTIENT RULES

\[
\frac{d}{dx}[f(x)f(x)] = f(x)\frac{d}{dx}[f(x)] + f(x)\frac{d}{dx}[f(x)] = 2f(x)\frac{d}{dx}[f(x)].
\]

For \( n = 3 \), \( g(x) = [f(x)]^3 \). Again, using the Product Rule, we have

\[
g'(x) = \frac{d}{dx}\{[f(x)]^3\} = \frac{d}{dx}\{[f(x)]^2f(x)\} = f(x)\frac{d}{dx}\{[f(x)]^2\} + [f(x)]^2\frac{d}{dx}[f(x)] = f(x)\left(2f(x)\frac{d}{dx}[f(x)]\right) + [f(x)]^2\frac{d}{dx}[f(x)] = 3[f(x)]^2\frac{d}{dx}[f(x)].
\]

We see that differentiation of \([f(x)]^n\) follows the same pattern as differentiation of \(x^n\),

\[
\frac{d}{dx}\{[f(x)]^n\} = n[f(x)]^{n-1}\frac{d}{dx}[f(x)]. \tag{2.71}
\]

This is in fact true for any real number \( n \). This can be proven for non-negative integers using the same approach as in the optional subsection on mathematical induction. It can be proven for negative integers using the differentiation rule for the reciprocal of a function; the general case, where \( n \) is not an integer, cannot be proven without concepts discussed in the next chapter. We now formally state the rule, which we accept without proof.

**Theorem 2.8 (General Power Rule)** If \( f \) is differentiable at \( x \), then for any real number \( n \),

\[
\frac{d}{dx}\{[f(x)]^n\} = n[f(x)]^{n-1}\frac{d}{dx}[f(x)], \tag{2.72}
\]

provided that \( f(x) \neq 0 \) if \( n < 0 \).

With these rules, and the rules presented earlier in this section, we can now compute the derivative of any polynomial, rational or algebraic function.
Example 2.19 We will use differentiation rules to compute \( f'(x) \), where
\[
f(x) = \frac{x}{\sqrt{x^2 + 1}}.  \tag{2.73}
\]
Using the Quotient Rule and the General Power Rule, we obtain
\[
f'(x) = \frac{d}{dx} \left[ \frac{x}{\sqrt{x^2 + 1}} \right]
= \frac{\sqrt{x^2 + 1} \frac{d}{dx} x - x \frac{d}{dx} \sqrt{x^2 + 1}}{[\sqrt{x^2 + 1}]^2}
= \frac{\sqrt{x^2 + 1} \cdot 1 - x \frac{d}{dx} [(x^2 + 1)^{1/2}]}{x^2 + 1}
= \frac{\sqrt{x^2 + 1} - x \frac{1}{2} (x^2 + 1)^{-1/2} \frac{d}{dx} [x^2 + 1]}{x^2 + 1}
= \frac{\sqrt{x^2 + 1} - x \frac{1}{2} (x^2 + 1)^{-1/2} (2x)}{x^2 + 1}
= \frac{\sqrt{x^2 + 1} - x (x^2 + 1)^{-1/2} (x)}{x^2 + 1}
= \frac{\sqrt{x^2 + 1} - x^2 (x^2 + 1)^{-1/2}}{x^2 + 1}
= \frac{\sqrt{x^2 + 1} - x^2 (x^2 + 1)^{-1/2} \sqrt{x^2 + 1}}{x^2 + 1}
= \frac{(x^2 + 1) - x^2}{(x^2 + 1)^{3/2}}
= \frac{1}{(x^2 + 1)^{3/2}}.
\]

\[\square\]

2.4.3 Other Trigonometric Functions

The derivatives of the remaining basic trigonometric functions, \( \tan x \), \( \cot x \), \( \sec x \), and \( \csc x \) can easily be obtained using the differentiation rules for \( \sin x \) and \( \cos x \), in conjunction with the Quotient Rule. For instance,
\[
\frac{d}{dx} [\tan x] = \frac{d}{dx} \left[ \frac{\sin x}{\cos x} \right]
= \frac{\cos x \frac{d}{dx} [\sin x] - \sin x \frac{d}{dx} [\cos x]}{\cos^2 x}
= \frac{\cos^2 x - \sin^2 x}{\cos^2 x}
= \sec^2 x.
\]
2.4. **THE PRODUCT AND QUOTIENT RULES**

\[
\begin{align*}
&= \cos x(\cos x) - \sin x(-\sin x) \\
&= \cos^2 x + \sin^2 x \\
&= \frac{1}{\cos^2 x} \\
&= \sec^2 x.
\end{align*}
\]

In summary, we have the differentiation rule

\[
\frac{d}{dx}[\tan x] = \sec^2 x. \tag{2.74}
\]

Using a similar approach, we can easily obtain the following additional rules,

\[
\frac{d}{dx}[\cot x] = -\csc^2 x, \quad \frac{d}{dx}[\sec x] = \sec x \tan x, \quad \frac{d}{dx}[\csc x] = -\csc x \cot x. \tag{2.75}
\]

We now illustrate the use of the six trigonometric differentiation rules that we have developed.

**Example 2.20** Let \( f(x) = \sec x \tan x \). Using the Product Rule, we obtain

\[
\begin{align*}
\frac{d}{dx}[f(x)] &= \frac{d}{dx}[\sec x \tan x] \\
&= \sec x \tan x \frac{d}{dx}[\sec x] + \sec x \tan x \frac{d}{dx}[\tan x] \\
&= \sec x (\sec x \tan x) + \sec x (\sec^2 x) \\
&= \sec x \tan^2 x + \sec^3 x. \tag{2.76}
\end{align*}
\]

Using the trigonometric identity \( \tan^2 x = \sec^2 x - 1 \), we can rewrite this derivative as

\[
\begin{align*}
\frac{d}{dx}[f(x)] &= \sec x \tan^2 x + \sec^3 x \\
&= \sec x (\sec^2 x - 1) + \sec^3 x \\
&= \sec^3 x - \sec x + \sec^3 x = 2\sec^3 x - \sec x. \tag{2.77}
\end{align*}
\]

\( \square \)

**Example 2.21** Let

\[
f(x) = \frac{\sin x}{x + \tan x}. \tag{2.78}
\]
Using the Quotient Rule, we obtain

\[
f'(x) = \frac{(x + \tan x) \frac{d}{dx}[\sin x] - \sin x \frac{d}{dx}[x + \tan x]}{(x + \tan x)^2}
\]

\[
= \frac{(x + \tan x)(\cos x) - \sin x(1 + \sec^2 x)}{(x + \tan x)^2}
\]

\[
= \frac{(x + \frac{\sin x}{\cos x}) (\cos x) - \sin x(1 + \sec^2 x)}{(x + \tan x)^2}
\]

\[
= \frac{x \cos x + \sin x - \sin x - \sin x \sec^2 x}{(x + \tan x)^2}
\]

\[
= \frac{x \cos x - \sin x \sec^2 x}{(x + \tan x)^2}
\]

\[\square\]

We have significantly expanded our range of functions that we can differentiate using the various rules we have constructed. However, these rules are not useful for various combinations of algebraic and trigonometric functions such as \( f(x) = \cos(x^2) \). To handle such combinations, which are \textit{compositions} of functions that can be differentiated using these rules, we will introduce yet another rule in the next section.

\textbf{Example 2.22} Differentiate

\[ f(x) = x \cos^2 x. \] (2.79)

\textbf{Solution} Using the Product Rule, in conjunction with the General Power Rule, we have

\[
f'(x) = \frac{d}{dx}[x \cos^2 x]
\]

\[
= \cos^2 x \frac{d}{dx}[x] + x \frac{d}{dx}[\cos^2 x]
\]

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

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

\[
= \cos^2 x - 2x \cos x \sin x.
\]
Example 2.23 Differentiate

\[ f(x) = \frac{1}{\sqrt{1 - \sin^2 x}}. \quad (2.80) \]

Solution We rewrite \( f(x) \) as \( f(x) = (1 - \sin^2 x)^{-1/2} \) and use the General Power Rule, twice, to obtain

\[
\begin{align*}
    f'(x) &= \frac{d}{dx}[(1 - \sin^2 x)^{-1/2}] \\
    &= -\frac{1}{2} (1 - \sin^2 x)^{-3/2} \frac{d}{dx}[1 - \sin^2 x] \\
    &= -\frac{1}{2} (1 - \sin^2 x)^{-3/2} \left(- \frac{d}{dx} [\sin^2 x] \right) \\
    &= \frac{1}{2} (1 - \sin^2 x)^{-3/2} \left(2 \sin x \frac{d}{dx} [\sin x] \right) \\
    &= \frac{1}{2} (1 - \sin^2 x)^{-3/2} (2 \sin x \cos x) \\
    &= (1 - \sin^2 x)^{-3/2} (\sin x \cos x) \\
    &= \sin x \cos x \\
    &= \frac{\sin x \cos x}{(1 - \sin^2 x)^{3/2}} \\
    &= \frac{\sin x \cos x}{\cos^2 x} \\
    &= \frac{\sin x}{\cos^2 x} \\
    &= \frac{1 \sin x}{\cos x \cos x} \\
    &= \sec x \tan x.
\end{align*}
\]

Recall that \( \sec x \tan x \) is the derivative of \( \sec x \). Simplifying our original function, we have

\[ f(x) = \frac{1}{\sqrt{1 - \sin^2 x}} = \frac{1}{\sqrt{\cos^2 x}} = \frac{1}{\cos x} = \sec x. \quad (2.81) \]

In simplifying both the function and its derivative, we have taken advantage of the trigonometric identity \( \sin^2 x + \cos^2 x = 1 \). \( \square \)
2.4.4 Rates of Change in the Natural and Social Sciences

The notion of the instantaneous rate of change of one quantity with respect to another plays a key role in many disciplines, and as a result researchers in these disciplines often need to use derivatives in their work. Here we give a brief list of some applications in which derivatives are useful.

- **Physics:** Many problems in physics deal with velocity, the rate of change of the position of an object with respect to time. Another key quantity that is a rate of change is *density*, which is the rate of change of the mass of an object with respect to its volume. In the simple case where an object is assumed to have constant density, we have the well-known formula, \( d = m/v \), where \( d \) denotes density, \( m \) denotes mass, and \( v \) denotes volume.

- **Chemistry:** The *rate of reaction* of a substance is defined to be the rate of change of the concentration of the substance with respect to time. In addition, the *compressibility* of a substance, such as a liquid or gas, is described by the expression
  \[
  \beta = -\frac{1}{V} \frac{dV}{dp}, \tag{2.82}
  \]
  where \( V \) denotes volume and \( p \) denotes pressure, under the assumption that the temperature of the substance is constant. Intuitively, as a substance is compressed, its volume decreases, so \( dV/dp < 0 \). The number \( \beta \) is therefore a positive number that indicates how rapidly, the volume of the substance decreases as it is compressed, per unit of volume.

- **Biology:** In biology, it is often necessary to study population growth or decline as determined by factors such as birth or death rates. In such cases, projections of future population are made using a determination of the rate of change of population with respect to time based on available data.

  Another application of the derivative is in the study of human anatomy. One example involves the rate of change of the velocity of blood as it travels through a blood vessel, with respect to the radius of the vessel.

- **Economics:** If a function \( C(x) \) denotes the cost of producing \( x \) units of a given product, then the *marginal cost* of production is defined by \( dC/dx \). Intuitively, this rate of change is an approximation of the
2.5. THE CHAIN RULE

additional cost of producing one more unit, given that \( x \) units are already being produced. The concepts of marginal demand, marginal revenue, and marginal profit are defined similarly.

Other applications of the derivative can be found in, for instance, geology, engineering, geography, meteorology, psychology, and sociology.

2.5 The Chain Rule

In the previous section, we pointed out that we could not use known differentiation rules to compute the derivative of various functions that are compositions of algebraic and/or trigonometric functions, such as \( h(x) = \sin(x^2) \). We can view this particular function as a composition of the functions \( f(x) = \sin x \) and \( g(x) = x^2 \). We say that \( h \) is the composition of \( f \) and \( g \), and write \( h = f \circ g \). The value of \( h(x) \) is given by the relation \( h(x) = f(g(x)) \).

In this section, we consider the problem of computing the derivative of a function \( h(x) \) that is a composition of two functions \( f \) and \( g \) whose derivatives are known. To solve this problem, we suppose that we have two functions \( f \) and \( g \) such that \( g \) is differentiable at \( a \) and \( f \) is differentiable at \( g(a) \). Our goal is to compute the derivative of \( f \circ g \) at \( a \), which we denote by \( (f \circ g)'(a) \). We begin with the simple relation

\[
f(g(a)+h) - f(g(a)) = f(g(a)+h) - f(g(a)) - hf'(g(a))
\]

which we rearrange algebraically to obtain

\[
f(g(a)+h) - f(g(a)) = hf'(g(a)) + h \left( \frac{f(g(a)+h) - f(g(a))}{h} - f'(g(a)) \right).
\]

We then define the function

\[
\epsilon_1(h) = \frac{f(g(a)+h) - f(g(a))}{h} - f'(g(a)).
\]

Intuitively, this function provides an indication of how well the difference quotient \( [f(g(a)+h) - f(g(a))]/h \) approximates the derivative of \( f \) at \( g(a) \), for any given \( h \).

From the definition of the derivative, and the fact that \( f \) is differentiable at \( g(a) \), it follows that

\[
\lim_{h \to 0} \epsilon_1(h) = 0.
\]
Therefore, if we define \( \epsilon_1(0) = 0 \), then \( \epsilon_1(h) \) is a continuous function of \( h \), and we have the relation

\[
f(g(a) + h) - f(g(a)) = h[f'(g(a)) + \epsilon_1(h)],
\]

which is valid for any number \( h \). Using a similar approach, we can obtain the relation

\[
g(a + h) - g(a) = h[g'(a) + \epsilon_2(h)],
\]

where \( \epsilon_2(h) \) is also a continuous function of \( h \), defined by

\[
\epsilon_2(h) = \begin{cases} \frac{g(a+h)-g(a)}{h} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}.
\]

Now, if we define

\[
\Delta g = g(a + h) - g(a),
\]

then

\[
f(g(a) + \Delta g) - f(g(a)) = \Delta g[f'(g(a)) + \epsilon_1(\Delta g)].
\]

From the definition of \( \Delta g \), we obtain

\[
f(g(a + h)) - f(g(a)) = h[g'(a) + \epsilon_2(h)][f'(g(a)) + \epsilon_1(\Delta g)].
\]

If we assume that \( h \neq 0 \), then we have

\[
\lim_{h \to 0} \frac{f(g(a + h)) - f(g(a))}{h} = [g'(a) + \epsilon_2(h)][f'(g(a)) + \epsilon_1(\Delta g)].
\]

Since \( g \) is differentiable at \( a \), it is also continuous at \( a \). It follows that the function \( g(a + h) \), viewed as a function of \( h \), is continuous at \( h = 0 \), and therefore

\[
\lim_{h \to 0} \Delta g = \lim_{h \to 0} [g(a + h) - g(a)] = \lim_{h \to 0} g(a + h) - \lim_{h \to 0} g(a) = g(a) - g(a) = 0.
\]

This implies that

\[
\lim_{h \to 0} \epsilon_1(\Delta g) = 0.
\]

Putting all of our results together, we obtain

\[
(f \circ g)'(a) = \lim_{h \to 0} \frac{f(g(a + h)) - f(g(a))}{h} = \lim_{h \to 0} [g'(a) + \epsilon_2(h)][f'(g(a)) + \epsilon_1(\Delta g)].
\]
2.5. **THE CHAIN RULE**

\[
\begin{align*}
\lim_{h \to 0} g'(a) + \epsilon_2(h) & \cdot \lim_{h \to 0} f'(g(a)) + \epsilon_1(\Delta g) \\
= [g'(a) + 0] \cdot [f'(g(a)) + 0] \\
= f'(g(a))g'(a).
\end{align*}
\]

We have just proved the following result.

**Theorem 2.9** (Chain Rule) If \( g \) is differentiable at \( x \) and \( f \) is differentiable at \( g(x) \), then the composition of \( f \) and \( g \), denoted by \( f \circ g \) and defined by \((f \circ g)(x) = f(g(x))\), is differentiable at \( x \) and

\[
\frac{d}{dx}[(f \circ g)(x)] = f'(g(x))g'(x). \quad (2.96)
\]

**Remark** Note that the derivative of the “outer” function in the composition, \( f \), is evaluated at \( g(x) \), while the derivative of the “inner” function, \( g \), is evaluated at \( x \). \( \square \)

This differentiation rule is called the Chain Rule because the derivatives of the two functions \( f \) and \( g \) that form the composition \( f \circ g \) are multiplied together to form the derivative of the composition; intuitively, the two derivatives are “chained” together. In general, a function that is a composition of \( n \) functions has a derivative that is the product of the derivatives of the \( n \) functions; informally, its derivative is said to be a “chain” of length \( n \).

**Example 2.24** Let \( h(x) = \sin(x^2) \). Then we can write \( h = f \circ g \), where \( f(x) = \sin x \) and \( g(x) = x^2 \). These two functions can be differentiated using rules introduced in previous sections, from which we obtain \( f'(x) = \cos x \) and \( g'(x) = 2x \). Applying the Chain Rule, we obtain

\[
h'(x) = \frac{d}{dx}[(f \circ g)(x)] = f'(g(x))g'(x) = \cos(g(x))g'(x) = \cos(x^2)(2x).
\]

\( \square \)

**Example 2.25** Let \( h(x) = \sqrt{\tan(x^2 + 1)} \). This function can be viewed as a composition of two functions, \( h = f \circ g \), where

\[
f(x) = \sqrt{x}, \quad g(x) = \tan(x^2 + 1).
\]

\( (2.98) \)
However, \( g(x) \) is itself a composition, of \( \tan x \) and \( x^2 + 1 \). Applying the Chain Rule to \( g(x) \), we obtain

\[
g'(x) = \tan(x^2 + 1) \sec^2 x. \tag{2.99}
\]

Using the Power Rule, we obtain

\[
f'(x) = \frac{d}{dx}[x^{1/2}] = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}. \tag{2.100}
\]

Applying the Chain Rule a second time yields

\[
h'(x) = \frac{1}{2\sqrt{\tan(x^2 + 1)}} \tan(x^2 + 1) \sec^2 x. \tag{2.101}
\]

In general, a composition of three functions \( f \circ g \circ h \) has the derivative

\[
\frac{d}{dx}[f \circ g \circ h] = \frac{d}{dx}[f \circ (g \circ h)] = f'(g(h(x))) \frac{d}{dx}[g \circ h] = f'(g(h(x)))g'(h(x))h'(x). \tag{2.103}
\]

In this case, the derivatives of \( f \), \( g \), and \( h \) are “chained together”, and evaluated at \( g(h(x)) \), \( h(x) \) and \( x \), respectively.

**Example 2.26** Recall the General Power Rule, introduced in Section 3.3,

\[
\frac{d}{dx}[(f(x))^n] = n[f(x)]^{n-1} \frac{d}{dx}[f(x)]. \tag{2.102}
\]

This rule can also be obtained using the Chain Rule. If we define \( h(x) = [f(x)]^n \), then \( h = g \circ f \), where \( g(x) = x^n \). Applying the Chain Rule yields

\[
\frac{d}{dx}[h(x)] = \frac{d}{dx}[(g \circ f)(x)] = g'(f(x))f'(x) = n[f(x)]^{n-1}f'(x). \tag{2.103}
\]

**Example 2.27** Let

\[
h(x) = \sqrt{\frac{x+1}{x-1}}. \tag{2.104}
\]
2.5. **THE CHAIN RULE**

Then \( h = f \circ g \), where
\[
f(x) = \sqrt{x}, \quad g(x) = \frac{x + 1}{x - 1}.
\]
(2.105)

Using the Quotient Rule, we obtain
\[
g'(x) = \frac{(x - 1) \frac{d}{dx}(x + 1) - (x + 1) \frac{d}{dx}(x - 1)}{(x - 1)^2} = \frac{(x - 1) \cdot 1 - (x + 1) \cdot 1}{(x - 1)^2} = -\frac{2}{(x - 1)^2}.
\]
(2.106)

It follows from the Chain Rule that
\[
h'(x) = f'(g(x))g'(x) = \frac{1}{2\sqrt{g(x)}}g'(x) = \frac{1}{2\sqrt{\frac{x + 1}{x - 1}}} \left(-\frac{2}{(x - 1)^2}\right) = -\frac{\sqrt{x - 1}}{x + 1} \frac{1}{(x - 1)^2}.
\]
(2.107)

**Example 2.28** Differentiate \( h(x) = \sin 2x \).

**Solution** The function \( h(x) \) can be viewed as a composition of two functions \( f(x) \) and \( g(x) \), where the **outer function** \( f(x) = \sin x \) and the **inner function** \( g(x) = 2x \). That is,
\[
h(x) = f(g(x)) = f(2x) = \sin(2x).
\]
(2.108)

From the Chain Rule,
\[
h'(x) = \frac{d}{dx}[f(g(x))] = f'(g(x))g'(x).
\]
(2.109)

Since
\[
f'(x) = \frac{d}{dx}[\sin x] = \cos x,
\]
(2.110)

it follows that
\[
f'(g(x)) = \cos(g(x)) = \cos(2x).
\]
(2.111)

Using the Chain Rule, we obtain
\[
h'(x) = \frac{d}{dx}[\sin(2x)]
= \frac{d}{dx}[f(g(x))]
= f'(g(x))g'(x)
= \cos(2x) \frac{d}{dx}[2x]
= \cos(2x) \cdot 2
= 2 \cos(2x).
\]
Example 2.29 Differentiate $h(x) = \cos(x^2)$.

Solution The function $h(x)$ can be viewed as a composition of two functions $f(x)$ and $g(x)$, where the outer function $f(x) = \cos x$ and the inner function $g(x) = x^2$. That is,

$$h(x) = f(g(x)) = f(x^2) = \cos(x^2). \quad (2.112)$$

From the Chain Rule,

$$h'(x) = \frac{d}{dx}[f(g(x))] = f'(g(x))g'(x). \quad (2.113)$$

Since

$$f'(x) = \frac{d}{dx}[\cos x] = -\sin x, \quad (2.114)$$

it follows that

$$f'(g(x)) = -\sin(g(x)) = -\sin(x^2). \quad (2.115)$$

Using the Chain Rule, we obtain

$$h'(x) = \frac{d}{dx}[\cos(x^2)]$$
$$= \frac{d}{dx}[f(g(x))]$$
$$= f'(g(x))g'(x)$$
$$= -\sin(x^2) \frac{d}{dx}[x^2]$$
$$= -\cos(x^2)(2x)$$
$$= -2x \cos(x^2).$$

\[
\]

Example 2.30 Differentiate $h(x) = \tan(\cos^2 x)$.

Solution The function $h(x)$ can be viewed as a composition of two functions $f(x)$ and $g(x)$, where the outer function $f(x) = \tan x$ and the inner function $g(x) = \cos^2 x$. That is,

$$h(x) = f(g(x)) = f(\cos^2 x) = \tan(\cos^2 x). \quad (2.116)$$
From the Chain Rule,

\[
h'(x) = \frac{d}{dx}[f(g(x))] = f'(g(x))g'(x). \tag{2.117}
\]

Since

\[
f'(x) = \frac{d}{dx}[	an x] = \sec^2 x, \tag{2.118}
\]

it follows that

\[
f'(g(x)) = \sec^2(g(x)) = \sec^2(\cos^2 x). \tag{2.119}
\]

To differentiate the inner function \( g(x) = \cos^2 x \), we can recognize that it, too, is a composition \( g(x) = r(s(x)) \), where the outer function \( r(x) = x^2 \) and the inner function \( s(x) = \cos x \). From the Chain Rule,

\[
g'(x) = \frac{d}{dx}[r(s(x))] = r'(s(x))s'(x). \tag{2.120}
\]

Since

\[
r'(x) = \frac{d}{dx}[x^2] = 2x, \tag{2.121}
\]

it follows that

\[
r'(s(x)) = 2s(x) = 2 \cos x. \tag{2.122}
\]

Finally, using the Chain Rule, we obtain

\[
h'(x) &= \frac{d}{dx}[\tan(\cos^2 x)] \\
      &= \frac{d}{dx}[f(g(x))] \\
      &= f'(g(x))g'(x) \\
      &= \sec^2(\cos^2 x) \frac{d}{dx}[\cos^2 x] \\
      &= \sec^2(\cos^2 x) \frac{d}{dx}[r(s(x))] \\
      &= \sec^2(\cos^2 x)r'(s(x))s'(x) \\
      &= \sec^2(\cos^2 x)(2 \cos x) \frac{d}{dx}[\cos x] \\
      &= \sec^2(\cos^2 x)(2 \cos x)(- \sin x) \\
      &= -2 \sec^2(\cos^2 x) \cos x \sin x.
\]

It should be noted that \( \sec^2(\cos^2 x) \) cannot be simplified using the relation \( \sec x = 1/\cos x \). It is important to realize that the expression \( \sec^2(\cos^2 x) \)
is not equivalent to \( \sec^2 x \cos^2 x \), which can be simplified to 1. In the first expression, we are taking the secant of \( \cos^2 x \) and squaring, whereas in the second, we are taking the secant of \( x \), squaring it, and multiplying by \( \cos^2 x \).

\[ \square \]

**Example 2.31** Differentiate \( h(x) = \sec \left( \frac{x+1}{x-1} \right) \sin((x^2 + 1)^3) \).

**Solution** First, we need to recognize that this function is a product of two functions, so we should begin by applying the Product Rule to obtain

\[
h'(x) = \frac{d}{dx} \left[ \sec \left( \frac{x+1}{x-1} \right) \sin((x^2 + 1)^3) \right]
\]

\[
= \sin((x^2 + 1)^3) \frac{d}{dx} \left[ \sec \left( \frac{x+1}{x-1} \right) \right] + \sec \left( \frac{x+1}{x-1} \right) \frac{d}{dx} \left[ \sin((x^2 + 1)^3) \right].
\]

We can now use the Chain Rule to compute the derivatives of two relatively simple functions. First, we consider

\[
h_1(x) = \sec \left( \frac{x+1}{x-1} \right).
\]

(2.123)

The function \( h_1(x) \) can be viewed as a composition of two functions \( f(x) \) and \( g(x) \), where the outer function \( f(x) = \sec x \) and the inner function \( g(x) = \frac{x+1}{x-1} \). That is,

\[
h_1(x) = f(g(x)) = f \left( \frac{x+1}{x-1} \right) = \sec \left( \frac{x+1}{x-1} \right).
\]

(2.124)

From the Chain Rule,

\[
h'_1(x) = \frac{d}{dx} [f(g(x))] = f'(g(x))g'(x).
\]

(2.125)

Since

\[
f'(x) = \frac{d}{dx} [\sec x] = \sec x \tan x,
\]

(2.126)

it follows that

\[
f'(g(x)) = \sec g(x) \tan g(x) = \sec \left( \frac{x+1}{x-1} \right) \tan \left( \frac{x+1}{x-1} \right).
\]

(2.127)

To differentiate the inner function \( g(x) \), we use the Quotient Rule to obtain

\[
g'(x) = \frac{d}{dx} \left[ \frac{x+1}{x-1} \right]
\]
2.5. THE CHAIN RULE

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

\[
= \frac{(x - 1)(1) - (x + 1)(1)}{(x - 1)^2}
\]

\[
= \frac{x - 1 - (x + 1)}{(x - 1)^2}
\]

\[
= \frac{x - 1 - x - 1}{(x - 1)^2}
\]

\[
= \frac{-2}{(x - 1)^2}.
\]

Using the Chain Rule, we obtain

\[
h'_1(x) = \frac{d}{dx} \left[ \sec \left( \frac{x + 1}{x - 1} \right) \right]
\]

\[
= \frac{d}{dx} [f(g(x))]
\]

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

\[
= \sec \left( \frac{x + 1}{x - 1} \right) \tan \left( \frac{x + 1}{x - 1} \right) \frac{d}{dx} \left[ \frac{x + 1}{x - 1} \right]
\]

\[
= \sec \left( \frac{x + 1}{x - 1} \right) \tan \left( \frac{x + 1}{x - 1} \right) \left( -\frac{2}{(x - 1)^2} \right)
\]

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

Next, we consider

\[
h_2(x) = \sin((x^2 + 1)^3).
\]

The function \( h_2(x) \) can be viewed as a composition of two functions \( f(x) \) and \( g(x) \), where the outer function \( f(x) = \sin x \) and the inner function \( g(x) = (x^2 + 1)^3 \). That is,

\[
h_2(x) = f(g(x)) = f((x^2 + 1)^3) = \sin ((x^2 + 1)^3).
\]

From the Chain Rule,

\[
h'_2(x) = \frac{d}{dx} [f(g(x))] = f'(g(x))g'(x).
\]

Since

\[
f'(x) = \frac{d}{dx} [\sin x] = \cos x,
\]

\[
(2.130)
\]

\[
(2.131)
\]
it follows that
\[ f'(g(x)) = \cos g(x) = \cos \left( (x^2 + 1)^3 \right). \tag{2.132} \]
To differentiate the inner function \( g(x) \), we use the General Power Rule to obtain
\[
g'(x) = \frac{d}{dx} [(x^2 + 1)^3]
= 3(x^2 + 1)^2 \frac{d}{dx} [x^2 + 1]
= 3(x^2 + 1)^2(2x)
= 6x(x^2 + 1)^2.
\]
Using the Chain Rule, we obtain
\[
h_2'(x) = \frac{d}{dx} [\sin((x^2 + 1)^3)]
= \frac{d}{dx} [f(g(x))]
= f'(g(x))g'(x)
= \cos \left( (x^2 + 1)^3 \right) \frac{d}{dx} [(x^2 + 1)^3]
= \cos \left( (x^2 + 1)^3 \right) [6x(x^2 + 1)^2]
= 6x \cos \left( (x^2 + 1)^3 \right) (x^2 + 1)^2.
\]
Finally, putting it all together, we have
\[
h'(x) = \frac{d}{dx} \left[ \sec \left( \frac{x+1}{x-1} \right) \sin((x^2 + 1)^3) \right]
= \sin((x^2 + 1)^3) \frac{d}{dx} \left[ \sec \left( \frac{x+1}{x-1} \right) \right] + \sec \left( \frac{x+1}{x-1} \right) \frac{d}{dx} \left[ \sin((x^2 + 1)^3) \right]
= \sin((x^2 + 1)^3) \left[ -2 \sec \left( \frac{x+1}{x-1} \right) \tan \left( \frac{x+1}{x-1} \right) \frac{1}{(x-1)^2} \right] +
\sec \left( \frac{x+1}{x-1} \right) [6x \cos \left( (x^2 + 1)^3 \right) (x^2 + 1)^2]
= \sec \left( \frac{x+1}{x-1} \right) \left\{ -2 \sin((x^2 + 1)^3) \tan \left( \frac{x+1}{x-1} \right) \frac{1}{(x-1)^2} +
6x \cos \left( (x^2 + 1)^3 \right) (x^2 + 1)^2 \right\}. \tag{2.133}
\]
2.6 Higher-Order Derivatives

Consider a function \( y = f(t) \), where \( t \) represents time and \( y \) represents position of an object along a line. We have discussed how the derivative \( f'(t) \) represents the instantaneous velocity of the object at time \( t \). However, just as velocity is the instantaneous rate of change of position with respect to time, acceleration is the instantaneous rate of change of velocity with respect to time. Therefore, the acceleration of the object at time \( t \) can be obtained by computing the derivative of \( f'(t) \). In other words, we can compute the acceleration by differentiating \( f \) twice. This leads to the following definition.

**Definition 2.8 (Second Derivative)** If a function \( f \) is differentiable at \( x \), and its derivative \( f' \) is also differentiable at \( x \), then the second derivative of \( f \) at \( x \), denoted \( f''(x) \), is

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

We also write

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

We say that \( f \) is **twice differentiable** at \( x \).

In Chapter 4, we will see that the second derivative is also useful in understanding the graph of a function, as well as finding minimum and maximum values of a function.

**Example 2.32** The height of a ball thrown into the air from a height of 40 ft with an initial velocity of 30 ft/s is given by the function \( f(t) = -16t^2 + 30t + 40 \), where \( t \) is the elapsed time in seconds since it was thrown. The velocity at time \( t \) is given by \( f'(t) = -32t + 30 \). The acceleration of the ball is given by the second derivative of \( f(t) \), denoted \( f''(t) \), which can be obtained by differentiating \( f'(t) \). This yields

\[
f''(t) = \frac{d^2}{dt^2}[f(t)] = \frac{d}{dt} \left( \frac{d}{dt}[f(t)] \right) = \frac{d}{dt}[-32t + 30] = -32.
\]

This second derivative can be interpreted as follows: the ball has constant acceleration of \(-32 \text{ ft/s}^2\), which is due to gravity. \( \square \)

It is natural to ask whether the rate of change of an object’s acceleration with respect to time is of interest. This quantity is called the **jerk** of the
object. The name comes from the fact that a rapid change in acceleration, which corresponds to a high rate of change in acceleration, tends to cause an abrupt motion in an object. Since the acceleration is given by the second derivative of a function \( f(t) \) that describes the position of an object, the jerk is obtained by differentiating \( f(t) \) three times. We therefore state the following definition.

**Definition 2.9** (Third Derivative) If a function \( f \) is twice differentiable at \( x \), and its second derivative \( f'' \) is also differentiable at \( x \), then the third derivative of \( f \) at \( x \), denoted \( f'''(x) \), is

\[
 f'''(x) = \frac{d}{dx}[f''(x)]. \tag{2.137}
\]

We also write

\[
 f'''(x) = \frac{d}{dx} \left( \frac{d^2}{dx^2} [f(x)] \right) = \frac{d^3}{dx^3} [f(x)]. \tag{2.138}
\]

We say that \( f \) is three times differentiable at \( x \).

**Example 2.33** Let

\[
 f(t) = 100 - \frac{1}{t + 0.01} \tag{2.139}
\]

be a function that describes the position of a car as it moves in a straight line, after the brakes are applied at \( t = 0 \) and the car is allowed to roll slowly to a stop, eventually stopping 100 feet after the brakes are first applied. The velocity of the car at time \( t \) is given by the first derivative,

\[
 f'(t) = \frac{d}{dt} \left[ 100 - \frac{1}{t + 0.01} \right] = \frac{d}{dt} [(t+0.01)^{-1}] = -(t+0.01)^{-2} \frac{d}{dt} [t+0.01] = -(t+0.01)^{-2}. \tag{2.140}
\]

The acceleration of the car at time \( t \) is given by the second derivative,

\[
 f''(t) = \frac{d^2}{dt^2} [f(t)] \\
 = \frac{d}{dt} [f'(t)] \\
 = \frac{d}{dt} [-(t+0.01)^{-2}] \\
 = -(2)(t+0.01)^{-3} \frac{d}{dt} [t + 0.01] \\
 = 2(t+0.01)^{-3}. \tag{2.141}
\]
Finally, the jerk at time \( t \) is given by the third derivative,
\[
f'''(t) = \frac{d^3}{dt^3} [f(t)]
\]
\[
= \frac{d}{dt} [f''(t)]
\]
\[
= \frac{d}{dt} [2(t + 0.01)^{-3}]
\]
\[
= -3(t + 0.01)^{-4} \frac{d}{dt} [t + 0.01]
\]
\[
= -3(t + 0.01)^{-4}.
\]
(2.142)

All three of these derivatives were obtained using the General Power Rule introduced in Section 3.3.

At \( t = 0.09 \), or 9/100ths of a second after the brakes are applied, the velocity of the car is given by
\[
f'(0.09) = -(0.09 + 0.01)^{-2} = -(0.1)^{-2} = -10^2 = -100 \text{ ft/s}.
\]
(2.143)

At the same time, the acceleration is
\[
f''(0.09) = 2(0.09 + 0.01)^{-3} = 2(0.1)^{-3} = 2(10^3) = 2,000 \text{ ft/s}^2.
\]
(2.144)

Finally, the jerk at \( t = 0 \) is
\[
f'''(0.09) = -3(0.09 + 0.01)^{-4} = -3(0.1)^{-4} = -3(10^4) = -30,000 \text{ ft/s}^3.
\]
(2.145)

It is reasonable to expect that the car experiences a jolt when the brakes are applied. The function \( f(t) \) is plotted in Figure 2.2. \( \square \)

In various applications it is necessary to work with functions obtained by differentiating a given function \( f(x) \) even more than three times; in fact, differentiating any number of times may be necessary. We define the order of a derivative of a function \( f \) to be the number of times that \( f \) must be differentiated to obtain the given derivative. For example, the order of \( f' \) is 1, the order of \( f'' \) is 2, and so on. We now state the definition of a derivative of arbitrary order.

**Definition 2.10 (General Higher-Order Derivatives)** If \( n \) is a nonnegative integer, then the \( n \)th derivative of a function \( f \) at \( x \), denoted by \( f^{(n)}(x) \), is
\[
f^{(n)}(x) = \begin{cases} 
  f(x) & \text{if } n = 0 \\
  \frac{d}{dx} [f^{(n-1)}(x)] & \text{if } n > 0
\end{cases}
\]
(2.146)
Figure 2.2: Motion of a car after applying its brakes at \( t = 0 \) and eventually coming to a stop 100 feet later. The point \( t = 0.09, f(0.09) = 90 \) at which the velocity, acceleration and jerk of the car are computed is indicated by the open circle.

provided that the \((n-1)st\) derivative of \( f \) at \( x \) exists. We also write

\[
  f^{(n)}(x) = \frac{d^n}{dx^n}[f(x)].
\]  

(2.147)

We say that \( f \) is \( n \) times differentiable at \( x \).

**Example 2.34** We compute the \( n \)th derivative of \( f(x) = x^k \), where \( k \geq n \). Assuming \( k \geq 3 \), we have

\[
  \frac{d}{dx}[x^k] = kn^{k-1}
\]

\[
  \frac{d^2}{dx^2}[x^k] = \frac{d}{dx}[kx^{k-1}] = k(k-1)x^{k-2}
\]

\[
  \frac{d^3}{dx^3}[x^k] = \frac{d}{dx}[k(k-1)x^{k-2}] = k(k-1)(k-2)x^{k-3}.
\]
From these derivatives, it is reasonable to conclude that
\[
\frac{d^n}{dx^n}[x^k] = k(k - 1) \cdots (k - n + 1)x^{k-n},
\]
and this is in fact the case, as can be proven by induction (see Section 3.3.1). This derivative can be expressed more concisely using the concept of the factorial of a number. The number \( n! \), read as “\( n \)-factorial,” is defined for any nonnegative integer \( n \) as follows:
\[
n! = \begin{cases} 
1 & \text{if } n = 0 \\
n \cdot (n - 1)! & \text{if } n > 0 
\end{cases}
\]
The factorial is actually defined for all real numbers but we do not need that definition in this book. Using the above definition, it can be shown that if \( n \) is a positive integer,
\[
n! = 1 \cdot 2 \cdot 3 \cdots n,
\]
that is, \( n! \) is the product of the first \( n \) positive integers. It can be seen that \( n! \) grows very rapidly as \( n \) increases: 2! = 2, 5! = 120, and 10! = 3,628,800.

Using this notation, we have
\[
\frac{d^n}{dx^n}[x^k] = k(k - 1) \cdots (k - n + 1)x^{k-n} = \frac{k(k - 1) \cdots (2)(1)}{(k - n) \cdots (2)(1)} x^{k-n} = \frac{k!}{(k - n)!}x^{k-n}.
\]
It follows that
\[
\frac{d}{dx^n}[x^n] = \frac{n!}{0!}x^{n-n} = n!
\]
Since \( n! \) is a constant, it also follows that
\[
\frac{d}{dx^n}[x^k] = 0
\]
whenever \( n > k \). □

**Example 2.35** Compute all of the derivatives of
\[
f(x) = x^4 - 4x^3 + 6x^2 - 4x + 1.
\]

**Solution** We use a combination of the following differentiation Rules: primarily, the Power Rule,
\[
\frac{d}{dx}[x^n] = nx^{n-1},
\]
the Sum and Difference Rules,
\[
\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x), \quad \frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)
\] (2.156)
so we can consider each term separately, and the Constant Multiple Rule,
\[
\frac{d}{dx}[cf(x)] = cf'(x).
\] (2.157)

Using these three rules yields the following derivatives:

\[
f'(x) = \frac{d}{dx}[x^4 - 4x^3 + 6x^2 - 4x + 1] \]
\[
= \frac{d}{dx}[x^4] - \frac{d}{dx}[4x^3] + \frac{d}{dx}[6x^2] - \frac{d}{dx}[4x] + \frac{d}{dx}[1] \]
\[
= \frac{d}{dx}[x^4] - 4 \frac{d}{dx}[x^3] + 6 \frac{d}{dx}[x^2] - 4 \frac{d}{dx}[x] + \frac{d}{dx}[1] \]
\[
= 4x^3 - 4(3x^2) + 6(2x) - 4(1) + 0 \]
\[
= 4x^3 - 12x^2 + 12x - 4
\]

\[
f''(x) = \frac{d}{dx}[4x^3 - 12x^2 + 12x - 4] \]
\[
= 4 \frac{d}{dx}[x^3] - 12 \frac{d}{dx}[x^2] + 12 \frac{d}{dx}[x] - \frac{d}{dx}[4] \]
\[
= 4(3x^2) - 12(2x) + 12(1) - 0 \]
\[
= 12x^2 - 24x + 12
\]

\[
f'''(x) = \frac{d}{dx}[12x^2 - 24x + 12] \]
\[
= 12 \frac{d}{dx}[x^2] - 24 \frac{d}{dx}[x] + \frac{d}{dx}[12] \]
\[
= 12(2x) - 24(1) + 0 \]
\[
= 24x - 24
\]

\[
f^{(4)}(x) = \frac{d}{dx}[24x - 24] \]
\[
= \frac{d}{dx}[24x] - \frac{d}{dx}[24] \]
\[
= 24 \frac{d}{dx}[x] - \frac{d}{dx}[24] \]
2.6. HIGHER-ORDER DERIVATIVES

\[ f(x) = 24(x) - 0 \]
\[ = 24. \]

Since the fourth derivative is a constant, it follows that \( f^{(n)}(x) = 0 \) for all \( n > 4 \).

If we examine the derivatives we just computed, we can observe the following:

\[
\begin{align*}
f(x) &= x^4 - 4x^3 + 6x^2 - 4x + 1 = (x - 1)^4 \\
f'(x) &= 4x^3 - 12x^2 + 12x - 4 = 4(x^3 - 3x^2 + 3x - 1) = 4(x - 1)^3 \\
f''(x) &= 12x^2 - 24x + 12 = 12(x^2 - 2x + 1) = 12(x - 1)^2 \\
f'''(x) &= 24x - 24 = 24(x - 1) \\
f^{(4)}(x) &= 24 \\
f^{(5)}(x) &= 0,
\end{align*}
\]

or,

\[
\begin{align*}
f(x) &= (x - 1)^4 \\
f'(x) &= 4(x - 1)^3 \\
f''(x) &= 4 \cdot 3(x - 1)^2 \\
f'''(x) &= 4 \cdot 3 \cdot 2(x - 1)^2 \\
f^{(4)}(x) &= 4 \cdot 3 \cdot 2 \cdot 1(x - 1)^0 \\
f^{(5)}(x) &= 4 \cdot 3 \cdot 2 \cdot 1 \cdot 0.
\end{align*}
\]

In general, the \( n \)th derivative of \([f(x)]^k\) has a constant factor that is equal to the product of all of the integers starting from \( k \) and counting down \( n \) times, where \( k \) itself counts as one time; that is, the product of all of the integers from \( k \) down to \( k - n + 1 \). Using the factorial notation,

\[ n! = 1 \cdot 2 \cdot \ldots \cdot n, \]  
(2.158)

we can express this constant factor concisely and obtain the following formula:

\[
\frac{d^n}{dx^n} ([f(x)]^k) = \frac{k!}{(k - n)!} [f(x)]^{k-n} f'(x). \]  
(2.159)

Example 2.36 Compute the first and second derivatives of

\[
f(x) = [\tan(x^3)]^{3/2}. \]  
(2.160)
Solution To compute the first derivative, we can recognize that \( f(x) \) is the composition of three functions, listed outermost first: \( x^{3/2} \), \( \tan x \), and \( x^3 \). We therefore use the Chain Rule, twice, to obtain

\[
f'(x) = \frac{d}{dx} \left\{ \left[ \tan(x^3) \right]^{3/2} \right\} = \frac{3}{2} \left[ \tan(x^3) \right]^{1/2} \frac{d}{dx} \left[ \tan(x^3) \right] = \frac{3}{2} \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) \frac{d}{dx} [x^3] = \frac{3}{2} \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3)(3x^2) = \frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3).
\]

To compute the second derivative, we need to recognize that the first derivative \( f'(x) \) is a product of three functions: \( \frac{9}{2} x^2 \), \( \left[ \tan(x^3) \right]^{1/2} \), and \( \sec^2(x^3) \). We therefore need to use the Product Rule. It is helpful to know that although the Product Rule is written for handling the product of two functions, it can easily be extended to handle a product of \( n \) factors as follows: your derivative consists of \( n \) terms, which each term including the derivative of one of the factors, times the other \( n - 1 \) factors. So, in this case, we have the following:

\[
f''(x) = \frac{d}{dx} \left\{ \frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) \right\} = \frac{d}{dx} \left\{ \frac{9}{2} x^2 \right\} \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) + \frac{9}{2} x^2 \frac{d}{dx} \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) + \frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \frac{d}{dx} [\sec^2(x^3)] = \left[ \frac{9}{2} (2x) \right] \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) + \frac{9}{2} x^2 \left[ \frac{1}{2} \right] \left[ \tan(x^3) \right]^{-1/2} \frac{d}{dx} [\tan(x^3)] \sec^2(x^3) + \frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \left[ 2 \sec(x^3) \frac{d}{dx} [x^3] \right] = \left[ \frac{9}{2} (2x) \right] \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) + \frac{9}{2} x^2 \left[ \frac{1}{2} \right] \left[ \tan(x^3) \right]^{-1/2} \sec^2(x^3) + \frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \cdot 2 \sec(x^3) \frac{d}{dx} [x^3]
\]
2.7. IMPLICIT DIFFERENTIATION

\[
\frac{9}{2} x^2 \left[ \frac{1}{2} \left( \tan(x^3) \right)^{-1/2} \sec^2(x^3)(3x^2) \right] \sec^2(x^3) + \\
\frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \left[ 2 \sec(x^3)(3x^2) \right]
\]

\[
= 9x \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) + \\
\frac{27}{4} x^4 \left[ \tan(x^3) \right]^{-1/2} \sec^4(x^3) + \\
27x^4 \left[ \tan(x^3) \right]^{1/2} \sec(x^3).
\]

An alternative approach would be to use the Product Rule in the way in which it originally written, but using it multiple times. For example, the differentiation process could begin as follows:

\[
f''(x) = \frac{d}{dx} \left\{ \frac{9}{2} x^2 \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) \right\}
\]

\[
= \frac{9}{2} x^2 \frac{d}{dx} \left\{ \left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) \right\} + \\
\left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) \frac{d}{dx} \left\{ \frac{9}{2} x^2 \right\}
\]

\[
= \frac{9}{2} x^2 \left[ \sec^2(x^3) \frac{d}{dx} \left\{ \left[ \tan(x^3) \right]^{1/2} \right\} \right] + \\
\left[ \tan(x^3) \right]^{1/2} \frac{d}{dx} \left\{ \sec^2(x^3) \right\} \right] + \\
\left[ \tan(x^3) \right]^{1/2} \sec^2(x^3) \frac{d}{dx} \left\{ \frac{9}{2} x^2 \right\}
\]

and then differentiation can continue in the same way as before. □

2.7 Implicit Differentiation

The central theme of this book is determining the instantaneous rate of change of one quantity, \( y \), with respect to another, \( x \). To this point, we have assumed that the relationship between \( x \) and \( y \) has been described by an equation of the form \( y = f(x) \). In this case, we are able to determine the instantaneous rate of change, which we denote by \( dy/dx \), by computing the derivative of \( f(x) \).

However, it is not always possible to describe the relationship between \( x \) and \( y \) by an equation of the form \( y = f(x) \), as we illustrate with an example.
Example 2.37 The unit circle, which is the circle of radius 1 with center at \((0,0)\), is described by the equation
\[
x^2 + y^2 = 1. \tag{2.161}
\]
It is not possible to rewrite this equation in the form \(y = f(x)\) to describe the entire circle; we can only describe either the top half or bottom half using the equation \(y = \pm \sqrt{1 - x^2}\). \(\square\)

In general, it is not possible to describe a curve in the \(xy\)-plane using an equation of the form \(y = f(x)\) whenever there is more than one \(y\)-value for any \(x\)-value, since the definition of a function requires that there is only one output for each input.

When an equation of the form \(y = f(x)\) is not available, we cannot obtain a function of \(x\) that describes the instantaneous rate of change \(dy/dx\); instead, we can obtain an equation that describes \(dy/dx\) in terms of \(x\) and \(y\). The process of obtaining such an equation is called implicit differentiation. This term arises from the fact that the equation relating \(x\) and \(y\) is called an implicit equation, since the function relating \(y\) and \(x\) is described implicitly by the equation, whereas an equation of the form \(y = f(x)\) is called an explicit equation, since the function relating \(y\) and \(x\) is described explicitly by the equation.

The process of implicit differentiation relies on the fact that if two functions of \(x\) and \(y\) are equal to one another, where \(y\) is understood to be a function of \(x\), then certainly their derivatives with respect to \(x\) must also be equal. Therefore, given an implicit equation relating \(x\) and \(y\), one can obtain an implicit equation describing \(dy/dx\) in terms of \(x\) and \(y\) by differentiating both sides of the original equation with respect to \(x\). In doing so, one must keep in mind that \(y\) is a function of \(x\), even if that function is not known; this leads to frequent usage of the Chain Rule.

Example 2.38 Consider the equation that describes the unit circle,
\[
x^2 + y^2 = 1. \tag{2.162}
\]
We differentiate both sides of this equation with respect to \(x\). Clearly, the right side of the equation, the constant 1, has a derivative of zero. As for the left side, differentiating \(x^2\) yields \(2x\), using the Power Rule. Finally, to differentiate \(y^2\) with respect to \(x\), we use the Power Rule again, in conjunction with the Chain Rule, to obtain
\[
\frac{d}{dx}[y^2] = 2y \frac{dy}{dx}. \tag{2.163}
\]
2.7. IMPLICIT DIFFERENTIATION

Putting all of these results together, we obtain an equation describing \( \frac{dy}{dx} \),

\[ 2x + 2y \frac{dy}{dx} = 0. \quad (2.164) \]

Solving for \( \frac{dy}{dx} \) yields

\[ \frac{dy}{dx} = -\frac{x}{y}, \quad y \neq 0. \quad (2.165) \]

This equation shows that \( y \) is not differentiable with respect to \( x \) at the points on the unit circle where \( y = 0 \). This makes sense, because at these points, which are \((-1, 0)\) and \((1, 0)\), the tangent line is vertical. \( \square \)

**Remark** Note that the equation describing \( \frac{dy}{dx} \) relates \( \frac{dy}{dx} \) to both \( x \) and \( y \). The dependence on \( y \) always occurs when differentiating implicitly, since otherwise one would have an explicit formula for \( \frac{dy}{dx} \) as a function of \( x \), which would imply that \( y \) could also be described explicitly as a function of \( x \). \( \square \)

**Example 2.39** We will compute \( \frac{dy}{dx} \), where

\[ \sin(xy) = \cos(x^2 + y). \quad (2.166) \]

We differentiate both sides implicitly with respect to \( x \) using the rules for differentiating trigonometric functions, along with the Chain Rule. This yields

\[ \cos(xy) \frac{d}{dx}[xy] = -\sin(x^2 + y) \frac{d}{dx}[x^2 + y]. \quad (2.167) \]

We then use the Product Rule on the left side and the Power and Sum Rules on the right side to obtain

\[ \cos(xy) \left( y + x \frac{dy}{dx} \right) = -\sin(x^2 + y) \left( 2x + \frac{dy}{dx} \right). \quad (2.168) \]

Rearranging terms to solve for \( \frac{dy}{dx} \) yields

\[ \frac{dy}{dx} = -\frac{2x \sin(x^2 + y) + y \cos(xy)}{x \cos(xy) + \sin(x^2 + y)}. \quad (2.169) \]

\( \square \)

Recall that an equation relating \( x \) and \( y \), whether implicit or explicit, describes a curve in the \( xy \)-plane. Once we have computed \( \frac{dy}{dx} \) in terms of \( x \) and \( y \), we can obtain the slope of the tangent line to the curve at any point \((x, y)\) on the curve by substituting the values of \( x \) and \( y \) into the equation for \( \frac{dy}{dx} \).
**Example 2.40** Consider the ellipse described by the equation

\[
\frac{x^2}{9} + \frac{y^2}{4} = 1. \tag{2.170}
\]

We wish to obtain the equation of the tangent line at the point \((1, \frac{4\sqrt{2}}{3})\). Differentiating the above equation implicitly with respect to \(x\) yields

\[
\frac{2x}{9} + \frac{1}{4} \frac{dy}{dx} = 0. \tag{2.171}
\]

Solving for \(\frac{dy}{dx}\) yields

\[
\frac{dy}{dx} = -\frac{4x}{9y}, \quad y \neq 0. \tag{2.172}
\]

Substituting \(x = 1\) and \(y = \frac{4\sqrt{2}}{3}\) into this equation yields the slope of the tangent line,

\[
\frac{dy}{dx} = -\frac{4 \cdot 1}{9(\frac{4\sqrt{2}}{3})} = -\frac{1}{3\sqrt{2}}. \tag{2.173}
\]

Since the tangent line passes through the point \((1, \frac{4\sqrt{2}}{3})\), we can obtain the equation of the tangent line using the point-slope form,

\[
y - \frac{4\sqrt{2}}{3} = -\frac{1}{3\sqrt{2}}(x - 1), \tag{2.174}
\]

which can be rearranged to obtain the slope-intercept form,

\[
y = -\frac{1}{3\sqrt{2}} x + \frac{1}{3\sqrt{2}} + \frac{4\sqrt{2}}{3}
\]

\[
= -\frac{1}{3\sqrt{2}} x + \frac{1}{3\sqrt{2}} + \frac{4\sqrt{2}}{3\sqrt{2}}
\]

\[
= -\frac{1}{3\sqrt{2}} x + \frac{1}{3\sqrt{2}} + \frac{8}{3\sqrt{2}}
\]

\[
= -\frac{1}{3\sqrt{2}} x + \frac{9}{3\sqrt{2}}
\]

\[
= -\frac{1}{3\sqrt{2}} x + \frac{3}{\sqrt{2}}
\]

The graph of the ellipse and the tangent line is shown in Figure 2.3. \(\square\)
Example 2.41 Consider the curve described by the equation
\[ \frac{x^2}{3} + \frac{y^2}{3} = 4. \] (2.175)
Compute the slope of the tangent line of this curve at the point \((-3\sqrt{3}, 1)\).

Solution This curve, which is called an astroid, is shown in Figure 2.4. To compute the slope of the tangent line at any point, we must first obtain an expression for \(dy/dx\). We can accomplish this by differentiating the equation of the astroid implicitly with respect to \(x\). Differentiating both sides of the equation \(x^{2/3} + y^{2/3} = 4\) yields
\[ \frac{d}{dx}[x^{2/3}] + \frac{d}{dx}[y^{2/3}] = \frac{d}{dx}[4] \] (2.176)
which, by the Power Rule and the Chain Rule, becomes
\[ \frac{2}{3} x^{-1/3} + \frac{2}{3} y^{-1/3} \frac{dy}{dx} = 0. \] (2.177)
Figure 2.4: The astroid $x^{2/3} + y^{2/3} = 4$, with its tangent line at the point $(-3\sqrt{3}, 1)$

We can now solve for $dy/dx$ by rearranging algebraically, which yields

$$\frac{2}{3} y^{-1/3} \frac{dy}{dx} = -\frac{2}{3} x^{-1/3}, \quad (2.178)$$

and then

$$\frac{dy}{dx} = -\frac{2}{3} \frac{x^{-1/3}}{y^{-1/3}} \quad (2.179)$$

which becomes

$$\frac{dy}{dx} = -\frac{x^{-1/3}}{y^{-1/3}} = -\frac{y^{1/3}}{x^{1/3}}. \quad (2.180)$$

To compute the slope of the tangent line at the point $(-3\sqrt{3}, 1)$, we can simply evaluate $dy/dx$ at this point by substituting $x = -3\sqrt{3}$ and $y = 1$.
into our expression for $dy/dx$. This yields

$$\frac{dy}{dx} = -\frac{1}{(\sqrt[3]{-3\sqrt{3}})^{1/3}} = -\frac{1}{\sqrt[3]{-3^{3/2}}} = \frac{1}{\sqrt[3]{3}},$$  \hspace{1cm} (2.181)$$

so the tangent line at this point has slope $1/\sqrt{3}$. The tangent line is shown in Figure 2.4. \qed

\section*{2.8 Related Rates}

Recall the Chain Rule for computing the derivative of a composition of functions $f \circ g$,

$$\frac{d}{dx}[(f \circ g)(x)] = \frac{d}{dx}[f(g(x))] = f'(g(x))g'(x).$$  \hspace{1cm} (2.182)$$

In this section, we discuss the interpretation of the Chain Rule, in the context of rates of change. Suppose that two quantities $u$ and $x$ are related by the equation $u = g(x)$, and that a third quantity $y$ is related to $u$ by the equation $y = f(u)$. Then, we have

$$\frac{dy}{du} = f'(u), \quad \text{and} \quad \frac{du}{dx} = g'(x).$$  \hspace{1cm} (2.183)$$

Since $y$ is related to $u$, and $u$ is related to $x$, it follows that $y$ is related to $x$ by the equation $y = f(g(x))$. Using the Chain Rule, we can describe the instantaneous rate of change of $y$ with respect to $x$, $dy/dx$, as follows:

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

(2.184)

We say that the rates of change $dy/dx$, $dy/du$ and $du/dx$ are all related rates with respect to one another, and that relationship is described by the Chain Rule.

This notion of related rates is useful in applications in which we have a number of quantities that are related to one another by known functions, and we need to determine the instantaneous rate of change of one particular quantity $y$ with respect to another quantity, $x$, even though we don’t have an explicit equation of the form $y = f(x)$ that relates them. In such cases, determining the rate of change using related rates can be easier than obtaining an explicit formula for the function $f$ that relates $x$ and $y$ and then computing $dy/dx = f'(x)$ directly, as we shall see in the following examples.
Example 2.42 Suppose that water is being poured at the constant rate of 100 cm$^3$/s into a bowl that is shaped like the lower half of a sphere of radius 30 cm. At what rate, in cm/s, is the height of the water in the bowl changing when the height is 10 cm?

Let $V$ denote the volume of water in the bowl. Then we have

$$\frac{dV}{dt} = 100 \text{ cm}^3/\text{s}. \quad (2.185)$$

The volume of water in the bowl is related to the height by the equation

$$V = \pi \left( 30h^2 - \frac{h^3}{3} \right). \quad (2.186)$$

Using integral calculus, one can derive formulas such as this one. In order to determine the rate of change of the height $h$ with respect to time, $dh/dt$, we can compute

$$\frac{dV}{dh} = \pi (60h - h^2) \quad (2.187)$$

and use the Chain Rule, which implies that

$$\frac{dV}{dt} = \frac{dV}{dh} \frac{dh}{dt}, \quad (2.188)$$

to obtain

$$\frac{dh}{dt} = \frac{dV}{dt} \frac{1}{\frac{dV}{dh}} = \frac{100}{\pi (60h - h^2)}. \quad (2.189)$$

Setting $h = 10$ yields the conclusion that when the height of water in the bowl is 10 cm, the height is changing at the rate of

$$\frac{100}{\pi (60(10) - 10^2)} = \frac{100}{500\pi} = \frac{5}{\pi} \text{ cm/s.} \quad (2.190)$$

Example 2.43 A lighthouse shines a beam of light that makes four revolutions per minute. During each revolution, the beam illuminates a straight shoreline. The lighthouse is 3 km away from the closest point on the shoreline, which we will denote by $P$. How fast is the beam moving along the shoreline when it is illuminating a point $Q$ that is 3 km along the shoreline from $P$?

Figure 2.5 illustrates the lighthouse and the points $P$ and $Q$ along the shoreline. We will assume that the beam is sweeping along the shoreline
in the counterclockwise direction, from an aerial view. Let \( y \) denote the position of the beam along the shoreline, with \( y = 0 \) corresponding to \( P \), and positive values of \( y \) corresponding to points that are “above” \( P \) in the figure, such as \( Q \). Let \( \theta \) denote the angle that the beam makes with the line between the lighthouse and \( P \). Then, as shown in the figure, \( \theta = \pi/4 \) radians when the beam is shining at the point \( Q \).

![Figure 2.5: Lighthouse beam illuminating a straight shoreline. The lighthouse is 3 km away from the closest point on the shoreline, indicated by \( P \). The point \( Q \) is located 3 km up the shoreline from \( P \).](image)

Since the lighthouse makes 4 revolutions per minute, we have

\[
\frac{d\theta}{dt} = 8\pi \text{ radians/min} = \frac{8\pi}{60} \text{ radians/s},
\]  

(2.191)

where \( t \) is measured in seconds. Furthermore, we can use right-triangle identities to obtain the relationship between \( y \) and \( \theta \),

\[
\tan \theta = \frac{y}{3} \implies y = 3\tan \theta,
\]

(2.192)
which yields
\[ \frac{dy}{d\theta} = 3 \sec^2 \theta \text{ km/radian.} \]  
(2.193)

From the Chain Rule, we have
\[ \frac{dy}{dt} = \frac{dy}{d\theta} \frac{d\theta}{dt}, \]  
(2.194)
which yields
\[ \frac{dy}{dt} = (3 \sec^2 \theta) \frac{8\pi}{60} = \frac{24\pi \sec^2 \theta}{60} = \frac{2\pi \sec^2 \theta}{5} \text{ km/s.} \]  
(2.195)

When the lighthouse is shining at the point \( Q \), \( \theta = \pi/4 \), so the speed at which the beam is traveling along the shoreline at that time is
\[ \frac{dy}{dt} = \frac{2\pi \sec^2(\pi/4)}{5} \]
\[ = \frac{2\pi}{5 \cos^2(\pi/4)} \]
\[ = \frac{2\pi}{5(\sqrt{2}/2)^2} \]
\[ = \frac{2\pi}{5/2} \]
\[ = \frac{4\pi}{5} \text{ km/s.} \]

\[ \square \]

**Example 2.44** A man is trying to escape from prison, and his escape route is along a long wall. A nearby guard tower is continuously shining a revolving spotlight around the grounds. The light makes 4 revolutions per minute, and the point \( P \) along the wall that is closest to the tower is 500 feet from the tower. How fast is the spotlight moving along the wall at a point \( Q \) that is 500 feet from \( P \)?

**Solution** A diagram of the problem is shown in Figure 2.6. We let \( y \) denote the position of the spotlight along the wall, where \( y = 0 \) corresponds to the point \( P \) and \( y = 500 \) corresponds to the point \( Q \), and let \( \theta \) denote the angle that the spotlight beam makes with the line between the tower and \( P \) (which, in Figure 2.6, corresponds to the \( x \)-axis). As usual, we let the variable \( t \) denote time. Using these variables, we can identify the quantity
that we seek, the speed at which the spotlight moves along the wall, as \( \frac{dy}{dt} \), at the point where \( y = 500 \) ft.

From the figure, we can see that \( y \) and \( \theta \) are related by the right-triangle identity

\[
\tan \theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{y}{500},
\]

which yields

\[
y = 500 \tan \theta.
\]

Differentiating both sides with respect to \( \theta \) yields

\[
\frac{dy}{d\theta} = 500 \sec^2 \theta.
\]

Since the spotlight makes four revolutions per minute, it follows that the angle \( \theta \) sweeps through a complete circle, which is \( 2\pi \) radians, four times
each minute. In other words,

\[ \frac{d\theta}{dt} = 8\pi \text{ radians/min} = \frac{8\pi}{60} \text{ radians/sec}. \tag{2.199} \]

By the Chain Rule,

\[
\frac{dy}{dt} = \frac{dy}{d\theta} \frac{d\theta}{dt} = 500 \sec^2 \theta \frac{8\pi}{60} = \frac{200\pi}{3} \sec^2 \theta.
\]

In order to determine \(dy/dt\) when \(y = 500\), we must determine the value of \(\theta\) at which \(y = 500\). Since \(y = 500 \tan \theta\), it follows that \(\tan \theta = 1\) when \(y = 500\), so \(\theta = \pi/4\). Alternatively, since we really want \(\sec \theta\), we can use the right-triangle identity and the information given in Figure 2.6 to obtain

\[
\sec \theta = \frac{\text{hypotenuse}}{\text{adjacent}} = \frac{\sqrt{500^2 + 500^2}}{500} = \frac{500\sqrt{2}}{500} = \sqrt{2}. \tag{2.200}
\]

This yields our final result,

\[
\frac{dy}{dt} = \frac{200\pi}{3} \sec^2 \theta = \frac{200\pi}{3} (\sec \theta)^2 = \frac{200\pi}{3} (\sqrt{2})^2 = \frac{400\pi}{3} \text{ ft/sec}.
\]

This example illustrates the typical approach to a related-rates problem. You are trying to compute a particular rate of change, based on knowledge of other rates of change that can be obtained by given information. By coming up with an equation that relates your unknown rate of change to your known rates of change, you can compute the unknown rate of change at a particular point, without having to determine explicit formulas that relate all of the quantities involved. ∎

**Example 2.45** Water is filling a cone-shaped container at a constant rate of 100 cm\(^3\)/s. The cone has a height of 20 cm and a radius of 10 cm. How
fast is the water level in the cone increasing when the cone is filled to a height of 10 cm?

**Solution** Let $V$ denote the volume of water in the cone at any time $t$, and let $h$ denote the height of the water (i.e., the water level) in the cone at time $t$. We want to compute $dh/dt$, the rate of change of the water level with respect to time.

We are given that

$$
\frac{dV}{dt} = 100 \text{ cm}^3/\text{s}. \quad (2.201)
$$

We need to somehow relate $dV/dt$ to $dh/dt$ to help us compute $dh/dt$. From the Chain Rule,

$$
\frac{dV}{dt} = \frac{dV}{dh} \frac{dh}{dt}, \quad (2.202)
$$

so we need to determine $dV/dh$. To accomplish this, we use the formula for the volume of a cone to obtain

$$
V = \frac{\pi}{3} r^2 h, \quad (2.203)
$$

where $r$ is the radius of the portion of the cone that is filled with water at time $t$. Since the radius of the cone is 10 cm and the height is 20 cm, it follows that $r = h/2$ at any particular time $t$. Therefore, the volume of water can be expressed as a function of the height of the water,

$$
V = \frac{\pi}{3} r^2 h = \frac{\pi}{3} \left( \frac{h}{2} \right)^2 h = \frac{\pi}{12} h^3, \quad (2.204)
$$

which yields

$$
\frac{dV}{dh} = \frac{\pi}{12} (3h^2) = \frac{\pi}{4} h^2. \quad (2.205)
$$

Now, we can determine how all of our rates of change are related. We have

$$
\frac{dV}{dt} = 100 = \frac{dV}{dh} \frac{dh}{dt} = \frac{\pi}{4} h^2 \frac{dh}{dt}, \quad (2.206)
$$

which can be rearranged to obtain

$$
\frac{dh}{dt} = \frac{100}{(\pi/4)h^2} = \frac{400}{\pi h^2}. \quad (2.207)
$$

To determine $dh/dt$ when the height of the water is 10 cm, we can simply substitute $h = 10$ and obtain

$$
\frac{dh}{dt} = \frac{400}{\pi(10^2)} = \frac{400}{100\pi} = \frac{4}{\pi} \text{ cm/s}. \quad (2.208)
$$
Chapter 2. Derivatives

Note that in this example, like the previous example, we identified the rate of change that we needed to compute, and used given information in order to determine related rates of change. We then used the relationship between all of these rates of change in order to obtain our answer. □

Example 2.46 A 10-foot ladder is leaning against a wall. If the end of the ladder that is on the floor is pulled away from the wall at a constant rate of 1 ft/s, how fast is the upper end of the ladder falling down when it is five feet above the floor?

Solution Let $x$ denote the horizontal distance along the floor, in feet, between the wall and the lower end of the ladder at any time $t$, and let $y$ denote the height of the upper end of the ladder above the floor, also in feet, at any time $t$. We wish to compute $\frac{dy}{dt}$ when $y = 5$. Since the ladder is ten feet long, we have

$$x^2 + y^2 = 10^2. \quad (2.209)$$

Since $x$ and $y$ both depend on the time $t$, we can differentiate this equation with respect to $t$ and obtain

$$2x \frac{dx}{dt} + 2y \frac{dy}{dt} = 0. \quad (2.210)$$

Solving for our unknown rate of change $\frac{dy}{dt}$ yields

$$\frac{dy}{dt} = \frac{-2x}{2y} \frac{dx}{dt} = -\frac{x}{y} \frac{dx}{dt}. \quad (2.211)$$

We now use the fact that $\frac{dx}{dt} = 1$ and $y = 5$ to obtain

$$\frac{dy}{dt} = -\frac{x}{5} \cdot 1 = -\frac{x}{5}. \quad (2.212)$$

From the relation $x^2 + y^2 = 100$, we have $x = \sqrt{100 - y^2} = \sqrt{100 - 5^2} = \sqrt{75}$, which yields

$$\frac{dy}{dt} = -\frac{\sqrt{75}}{5} = -\frac{5\sqrt{3}}{5} = -\sqrt{3} \text{ ft/s}. \quad (2.213)$$

That is, the upper end of the ladder is sliding down the wall at a rate of $\sqrt{3} \approx 1.732$ ft/s when it is 5 feet above the floor.

It is worthwhile to note that in this example, the rates of change were not related by the Chain Rule, but by the Pythagorean Theorem. Still, the general approach is the same: identify the rate of change that needs to be
2.8. RELATED RATES

computed, use given information to compute related rates of change, and then use the relationship among the rates of change to obtain the solution.

Example 2.47 Given a curve that is described by the equation

\[(x^2 + y)^2 = 10(x^2 - y^3),\]  
\[(2.214)\]

compute the equation of the tangent line to the curve at the point \((\sqrt{10}, 0)\).

**Solution** We begin by differentiating both sides of the equation that describes the curve with respect to \(x\). Differentiating the left side, using the Power Rule and the Chain Rule, yields

\[
\frac{d}{dx}[(x^2 + y)^2] = 2(x^2 + y) \frac{d}{dx}[x^2 + y] = 2(x^2 + y) \left(2x + \frac{dy}{dx}\right). \tag{2.215}
\]

To differentiate the right side, we use the Constant Multiple Rule, the Sum Rule, the Power Rule and the Chain Rule to obtain

\[
\frac{d}{dx}[10(x^2 - y^3)] = 10 \frac{d}{dx}[x^2 - y^3] = 10 \left(2x - 3y^2 \frac{dy}{dx}\right). \tag{2.216}
\]

Having differentiated both sides, we now have an equation that describes \(dy/dx\),

\[
2(x^2 + y) \left(2x + \frac{dy}{dx}\right) = 10 \left(2x - 3y^2 \frac{dy}{dx}\right). \tag{2.217}
\]

We now need to solve this equation for \(dy/dx\). Expanding both sides, we have

\[
2(x^2)(2x) + 2y(2x) + 2x^2 \frac{dy}{dx} + 2y \frac{dy}{dx} = 10(2x) - 10(3y^2) \frac{dy}{dx}, \tag{2.218}
\]

or

\[
4x^3 + 4xy + 2x^2 \frac{dy}{dx} + 2y \frac{dy}{dx} = 20x - 30y^2 \frac{dy}{dx}. \tag{2.219}
\]

We move all terms involving \(dy/dx\) to the left side, and all terms not involving \(dy/dx\) to the right side, to obtain

\[
2x^2 \frac{dy}{dx} + 2y \frac{dy}{dx} + 30y^2 \frac{dy}{dx} = -4x^3 - 4xy + 20x \tag{2.220}
\]

or

\[
[2x^2 + 2y + 30y^2] \frac{dy}{dx} = -4x^3 - 4xy + 20x. \tag{2.221}
\]
It follows that
\[
\frac{dy}{dx} = \frac{-4x^3 - 4xy + 20x}{2x^2 + 2y + 30y^2}.
\] (2.222)

To obtain the slope of the tangent line at the point \((\sqrt{10}, 0)\), we substitute \(\sqrt{10}\) for \(x\) and 0 for \(y\) and obtain
\[
\frac{dy}{dx} = \frac{-4(\sqrt{10})^3 + 20(\sqrt{10})}{2(\sqrt{10})^2} = \frac{-4(10)\sqrt{10} + 20\sqrt{10}}{2(10)} = \frac{-40\sqrt{10} + 20\sqrt{10}}{20} = -\sqrt{10}.
\] (2.223)

We can now obtain the equation of this tangent line using the point-slope form,
\[
y - y_0 = m(x - x_0),
\] (2.224)
where \(m\) is the slope, \(-\sqrt{10}\), and \((x_0, y_0)\) is the point of tangency, with \(x_0 = \sqrt{10}\) and \(y_0 = 0\). Substituting the values of \(m\), \(x_0\) and \(y_0\) yields the equation
\[
y = -\sqrt{10}(x - \sqrt{10}) = -\sqrt{10}x + 10.
\] (2.225)

Example 2.48 Suppose that the length of the side of a perfect cube is increasing at the rate of 5 cm/s. How fast is the volume of the cube changing when the side length is 10 cm?

Solution We follow the general approach to solving related rates problems:

1. We identify all relevant quantities in the problem. Let \(s\) denote the side length of the cube, \(V\) denote the volume of the cube, and \(t\) denote time. While time is not explicitly mentioned, it is implied by the mention of the rate of change of the side length, which is in cm/s.

2. We identify what exactly needs to be computed. The problem asks for the rate of change of the volume of the cube with respect to time, which is \(dV/dt\), when the side length is 10 cm; that is, \(s = 10\).

3. We identify relationships among the quantities in the problem from the problem statement. We are given that the side length, \(s\), is increasing at the rate of 5 cm/s, which implies that \(ds/dt = 5\). Furthermore, the side length \(s\) of the cube and the volume \(V\) of the cube are related by the formula for the volume of a cube; that is, \(V = s^3\).

4. We differentiate relationships among quantities implicitly with respect to the independent variable of our desired rate of change. Our goal is
2.8. RELATED RATES

to compute \( dV/dt \), and the independent variable in this rate of change is \( t \). Therefore, we differentiate relationships among the quantities in our problem with respect to \( t \). This will indicate how the various rates of change in our problem are actually related to one another, which, in turn, will enable us to compute the desired rate of change.

In this problem, we only have one such relationship to differentiate with respect to \( t \), \( V = s^3 \). We have

\[
\frac{dV}{dt} = \frac{d}{dt}[s^3] = 3s^2 \frac{ds}{dt}.
\] (2.226)

5. We use given information to compute the desired rate of change at the indicated point. We can see that our desired rate of change, \( dV/dt \), can be obtained from our knowledge of \( ds/dt \) and \( s \). We are given that \( ds/dt = 5 \), and are asked to compute \( dV/dt \) when \( s = 10 \). We have

\[
\frac{dV}{dt} = 3s^2 \frac{ds}{dt} = 3(10^2)(5) = 1500 \text{ cm}^3/\text{s},
\] (2.227)

and our solution is complete.

\[\square\]

Example 2.49 The volume of the box is changing at a rate of 100 cm\(^3\)/s, while the length remains constant and the width is changing at a rate of 4 cm/s. How fast is the height of the box changing when the volume is 2000 cm\(^3\), the width is 8 cm, and the height is 20 cm?

Solution As in the previous example, we follow the general approach to solving related rates problems:

1. We identify all relevant quantities in the problem. Let \( \ell \) denote the length of the box, \( w \) denote its width, \( h \) denote its height, and \( V \) denote its volume. Again, we let \( t \) denote time. While time is not explicitly mentioned, it is implied by the mention of the rate of change of the width, which is in cm/s.

2. We identify what exactly needs to be computed. The problem asks for the rate of change of the height of the box with respect to time, which is \( dh/dt \), when \( V = 2000 \), \( w = 8 \), and \( h = 20 \).

3. We identify relationships among the quantities in the problem from the problem statement. We are given that the volume, \( V \), is increasing at
the rate of 100 cm$^3$/s, which implies that $dV/dt = 100$. Furthermore, the length, $\ell$, remains constant, so we have $d\ell/dt = 0$. In addition, the width, $w$ is changing at the rate of 4 cm/s, so $dw/dt = 4$. Finally, the dimensions $\ell$, $w$ and $h$ of the box and the volume $V$ of the box are related by the formula for the volume of a box; that is, $V = \ell wh$.

4. We differentiate relationships among quantities implicitly with respect to the independent variable of our desired rate of change. Our goal is to compute $dh/dt$, and the independent variable in this rate of change is $t$. Therefore, we differentiate relationships among the quantities in our problem with respect to $t$. This will indicate how the various rates of change in our problem are actually related to one another, which, in turn, will enable us to compute the desired rate of change.

In this problem, we only have one such relationship to differentiate with respect to $t$, $V = \ell wh$. We have, by the Product Rule,

$$
\frac{dV}{dt} = \frac{d}{dt}[\ell wh]
$$

$$
= \frac{d}{dt}[\ell(wh)]
$$

$$
= wh\frac{d\ell}{dt} + \ell \frac{d}{dt}[wh]
$$

$$
= wh\frac{d\ell}{dt} + \ell \left[ h \frac{dw}{dt} + w \frac{dh}{dt} \right]
$$

$$
= wh\frac{d\ell}{dt} + \ell h \frac{dw}{dt} + \ell w \frac{dh}{dt}.
$$

5. We use given information to compute the desired rate of change at the indicated point. We can see that our desired rate of change, $dh/dt$, can be obtained from our knowledge of $dV/dt$, $dw/dt$, $d\ell/dt$, $V$, $w$ and $h$. We are given that $dV/dt = 100$, $d\ell/dt = 0$, and $dw/dt = 4$, and are asked to compute $dh/dt$ when $V = 2000$, $w = 8$, and $h = 20$. From

$$
\frac{dV}{dt} = wh\frac{d\ell}{dt} + \ell h \frac{dw}{dt} + \ell w \frac{dh}{dt}
$$

(2.228)

we have

$$
100 = 8(20)(0) + \ell(20)(4) + \ell(8) \frac{dh}{dt},
$$

(2.229)

which simplifies to

$$
100 = 80\ell + 8\ell \frac{dh}{dt}.
$$

(2.230)
We are not given \( \ell \), but we can obtain \( \ell \) from the formula \( V = \ell wh \) and the fact that \( V = 2000 \), \( w = 8 \) and \( h = 20 \). This yields

\[
\ell = \frac{V}{wh} = \frac{2000}{8(20)} = \frac{100}{8} = 12.5 \text{ cm.} \tag{2.231}
\]

It follows that

\[
100 = 80(12.5) + 8(12.5) \frac{dh}{dt}, \tag{2.232}
\]

which becomes

\[
100 = 1000 + 100 \frac{dh}{dt} \tag{2.233}
\]

and therefore

\[
\frac{dh}{dt} = \frac{100 - 1000}{100} = \frac{-900}{100} = -9 \text{ cm/s}, \tag{2.234}
\]

and our solution is complete.

\[\Box\]

**Example 2.50** Suppose that \( y = \cos x \) and \( z = (x^2 + 1)^2 \). Compute \( dy/dz \) when \( x = 3 \).

**Solution** As in the previous example, we follow the general approach to solving related rates problems:

1. *We identify all relevant quantities in the problem.* In this problem, the quantities are already identified; they are \( x \), \( y \) and \( z \).

2. *We identify what exactly needs to be computed.* This is already explicitly prescribed in the problem statement; we are to compute \( dy/dz \) when \( x = 3 \).

3. *We identify relationships among the quantities in the problem from the problem statement.* These relationships are already given in the problem statement; \( x \) and \( y \) are related by the equation \( y = \cos x \), and \( x \) and \( z \) are related by the equation \( z = (x^2 + 1)^2 \).

4. *We differentiate relationships among quantities implicitly with respect to the independent variable of our desired rate of change.* Our goal is to compute \( dy/dz \), and the independent variable in this rate of change is \( z \). Therefore, we differentiate relationships among the quantities in our problem with respect to \( z \). This will indicate how the various rates
of change in our problem are actually related to one another, which, in turn, will enable us to compute the desired rate of change.

In this problem, we have two such relationships to differentiate with respect to $z$, $y = \cos x$ and $z = (x^2 + 1)^2$. Differentiating $y = \cos x$ implicitly with respect to $z$, by the Chain Rule,

$$\frac{dy}{dz} = \frac{d}{dz} \cos x = -\sin x \frac{dx}{dz}. \quad (2.235)$$

The left side of this new equation has our desired rate of change, $dy/dz$, but we have no information about $dx/dz$. We can obtain information about this rate of change by differentiating our other known relationship, $z = (x^2 + 1)^2$, with respect to $z$. Using the Power Rule and the Chain Rule, we obtain

$$\frac{dz}{dz} = \frac{d}{dz} [(x^2 + 1)^2] = 2(x^2 + 1) \frac{d}{dz} [x^2 + 1] = 2(x^2 + 1)(2x) \frac{dx}{dz}. \quad (2.236)$$

Simplifying yields the equation

$$1 = 4x(x^2 + 1) \frac{dx}{dz}, \quad (2.237)$$

or

$$\frac{dx}{dz} = \frac{1}{4x(x^2 + 1)}. \quad (2.238)$$

5. We use given information to compute the desired rate of change at the indicated point. We can see that our desired rate of change, $dy/dz$, can be obtained from our knowledge of $dx/dz$ and $x$. We are asked to compute $dy/dz$ when $x = 3$. From

$$\frac{dy}{dz} = -\sin x \frac{dx}{dz} \quad (2.239)$$

and

$$\frac{dx}{dz} = \frac{1}{4x(x^2 + 1)} \quad (2.240)$$

we have

$$\frac{dy}{dz} = -\sin x \frac{1}{4x(x^2 + 1)}. \quad (2.241)$$

Substituting $x = 3$ yields

$$\frac{dy}{dz} = -\frac{\sin 3}{4(3)(3^2 + 1)} = -\frac{\sin 3}{120} \approx -0.001176, \quad (2.242)$$

and our solution is complete.
It is worthwhile to note that we can differentiate the relationships given in the problem statement with respect to $x$ and obtain

\[
\frac{dy}{dx} = -\sin x, \quad \frac{dz}{dx} = 4x(x^2 + 1). \tag{2.243}
\]

Earlier, we found that

\[
\frac{dy}{dz} = -\frac{\sin x}{4x(x^2 + 1)} \tag{2.244}
\]

which implies

\[
\frac{dy}{dz} = \frac{dy}{dx} \cdot \frac{dz}{dx}, \tag{2.245}
\]

which can also be obtained directly from the Chain Rule, which states that

\[
\frac{dy}{dx} = \frac{dy}{dz} \cdot \frac{dz}{dx}. \tag{2.246}
\]
Chapter 3

Inverse Functions

3.1 Exponential Functions

In many applications, the relationship between two quantities \( x \) and \( y \) is described by an equation of the form \( y = f(x) \), where \( f \) is a function, and it is known that the rate of change of \( y \) with respect to \( x \) is proportional to the value of \( y \). When this is the case, the function \( f(x) \) is an exponential function. In this section, we begin our discussion of exponential functions and learn about some applications in which they are useful. We begin with a precise definition of an exponential function.

**Definition 3.1** (Exponential function) Let \( b \) be a positive real number that is not equal to 1. The exponential function with base \( b \) is the function \( f(x) \) that is defined by

\[
 f(x) = b^x \tag{3.1}
\]

where \( x \) is any real number. The number \( b \) is called the base of the function \( f(x) \), and the number \( x \) is called the exponent.

Certainly, the most familiar exponential function is the exponential function with base 10, \( f(x) = 10^x \). This function is used in scientific notation, which is used to express numbers in such a way that its magnitude and its significant digits are separated and therefore more easily read. For example, the number 123,456.789 is written as \( 1.23456789 \times 10^5 \), while 0.000001234 is written as \( 1.234 \times 10^{-6} \). Another well-known exponential function is the one with base 2, \( f(x) = 2^x \). Various powers of 2 are often used in computer science.

Before continuing, we recall certain basic properties of exponents.
• \( b^0 = 1 \) for any real number \( b \) (including zero!)

• If \( x \) is a positive integer, then \( b^x \) is the product of \( b \times b \times \cdots \times b \), where \( b \) appears as a factor \( x \) times. For example, \( b^1 = b \), \( b^2 = b \times b \), and so on.

• If \( x \) is a negative integer, then \( b^x = (1/b)^{-x} \).

• If \( x \) is a rational number, then \( x = p/q \), where \( p \) and \( q \) are integers, and \( b^x \) is equal to the \( q \)th root of \( b^p \).

If \( x \) is an irrational number, such as \( \sqrt{2} \), we can define \( b^x \) as a limit of the sequence \( b^{x_1}, b^{x_2}, \) and so on, where \( x_1, x_2, \ldots \) is a sequence of numbers that converges to \( x \). We will provide a more practical definition in an upcoming section.

**Example 3.1** Since 4 is a positive integer, \( 2^4 \) is simply the product of four 2’s. That is, \( 2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16 \). □

**Example 3.2** Since \(-3\) is a negative integer, \( 10^{-3} \) is the reciprocal of \( 10^3 \). That is,

\[
10^{-3} = \frac{1}{10^3} = \frac{1}{10 \cdot 10 \cdot 10} = \frac{1}{1,000} = 0.001.
\]

□

**Example 3.3** The expression \( 25^{1/2} \) is the “second root”, or square root, of the base 25, so \( 25^{1/2} = \sqrt{25} = 5 \). □

**Example 3.4** The expression \( 8^{2/3} \) is the square of the cube root of 8. That is,

\[
8^{2/3} = (8^{1/3})^2 = 2^2 = 4.
\]

□

### 3.1.1 Basic Exponential Graphs

The graphs of two exponential functions, with bases 2 and 1/3, are shown in Figure 3.1. From these two graphs, we make the following observations, which apply to exponential functions in general.

• The graph of any exponential function contains the point \((0, 1)\). This is due to the fact that any number raised to the zeroth power is equal to 1.
3.1. EXPONENTIAL FUNCTIONS

Figure 3.1: Graphs of \( y = 2^x \) (left plot) and \( y = (1/3)^x \) (right plot). In each plot, the point \((0, 1)\) is indicated by a circle.

- The graph is a continuous curve that does not have any breaks or sharp corners in it.
- The graph of \( y = b^x \) has a horizontal asymptote at \( y = 0 \). Specifically, the graph approaches the horizontal line \( y = 0 \) as \( x \) approaches \(-\infty\) if \( b > 1 \), and it approaches this same line as \( x \) approaches \( \infty \) if \( 0 < b < 1 \).
- If \( b > 1 \), then \( b^x \) is increasing. This means that if \( x > y \), then \( b^x > b^y \). On the other hand, if \( 0 < b < 1 \), then \( b^x \) is decreasing, meaning that \( x > y \) implies that \( b^x < b^y \).
- The function \( b^x \) is one-to-one, which means that if \( b^x = b^y \), then we must have \( x = y \). Equivalently, \( b^x \) passes the horizontal line test, meaning that any horizontal line passes through the graph of \( b^x \) at most once. This implies that \( b^x \) has an inverse function. The inverse
function of an exponential function is called a logarithmic function; we will study these functions in upcoming sections.

Some of these properties can be described more concisely in terms of limits. Specifically, if \( b > 1 \), then we have

\[
\lim_{x \to \infty} b^x = \infty, \quad \lim_{x \to -\infty} b^x = 0,
\]

whereas if \( b < 1 \), then

\[
\lim_{x \to \infty} b^x = 0, \quad \lim_{x \to -\infty} b^x = \infty.
\]

**Example 3.5** The graphs of \( y = 2^x \) and \( y = 3^x \) are shown in Figure 3.2. Note that as \( x \) increases toward \( \infty \), \( 3^x \) increases more rapidly than \( 2^x \). Also, as \( x \) decreases toward \( -\infty \), \( 3^x \) decreases toward zero more rapidly than \( 2^x \). Finally, note that the two graphs intersect at the point \((0, 1)\), since this point is included in the graph of *every* exponential function of the form \( b^x \).

**Example 3.6** We will compute

\[
\lim_{x \to 0^+} 2^{1/x}.
\]

From

\[
\lim_{x \to 0^+} \frac{1}{x} = \infty,
\]

it follows that if we let \( t = 1/x \), we have

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

### 3.1.2 Additional Exponential Properties

Exponential functions obey the following *laws of exponents*. These laws are likely familiar in the case of rational exponents, but they actually apply for all real exponents. As before, we assume that the bases, in this case \( a \) and \( b \), are positive, and not equal to 1.

\[
\begin{align*}
\frac{a^x}{a^y} &= a^{x-y} & (3.6) \\
(a^x)^y &= a^{xy} & (3.3) \\
(ab)^x &= a^x b^x & (3.4) \\
\left(\frac{a}{b}\right)^x &= \frac{a^x}{b^x} & (3.5) \\
(a^x)^{1/y} &= a^{x/y} & (3.2)
\end{align*}
\]
3.1. EXPONENTIAL FUNCTIONS

Figure 3.2: Graphs of \( y = 2^x \) (solid curve) and \( y = 3^x \) (dashed curve)

These properties can be helpful in computing limits involving exponential functions.

**Example 3.7** We will compute

\[
\lim_{x \to \infty} \frac{10^x}{10^x + 1}.
\]

If we divide the numerator and denominator by \(10^x\), we obtain

\[
\lim_{x \to \infty} \frac{10^x}{10^x + 1} = \lim_{x \to \infty} \frac{10^x}{10^x + 1} \frac{1/10^x}{1/10^x} = \lim_{x \to \infty} \frac{10^x/10^x}{(10^x + 1)/10^x} = \lim_{x \to \infty} \frac{10^x - x}{10^x/10^x + 1/10^x}
\]
\[
\begin{align*}
&= \lim_{x \to \infty} \frac{10^0}{10^0 + 10^0/10^x} \\
&= \lim_{x \to \infty} \frac{1}{1 + 10^{-x}} \\
&= \frac{1}{1 + 0} \\
&= 1.
\end{align*}
\]

We have used the fact that
\[
\lim_{x \to \infty} 10^{-x} = \lim_{t \to -\infty} 10^t = 0,
\]
where \( t = -x \).

In addition to the above laws of exponents, exponential functions satisfy these properties that are particularly useful in solving equations involving exponents: first, if \( x \neq 0 \), then
\[
a^x = b^x \quad \text{if and only if} \quad a = b. \tag{3.7}
\]

This property allows one to perform certain common operations on equations to obtain equivalent equations, such as squaring both sides or taking the square root of both sides. Second, we have
\[
a^x = a^y \quad \text{if and only if} \quad x = y. \tag{3.8}
\]

Using this property, one can solve an equation involving exponents by expressing both sides in terms of the same base.

**Example 3.8** Since \( a^x a^y = a^{x+y} \), it follows that \( 2^2 2^3 = 2^{2+3} = 2^5 = 32 \). We can verify this by noting that \( 2^2 2^3 = 4 \cdot 8 = 32 \). □

**Example 3.9** Since \( (a^x)^y = a^{xy} \), it follows that \( (2^3)^2 = 2^6 = 64 \). We can verify this by noting that \( (2^3)^2 = 8^2 = 64 \). □

**Example 3.10** We will simplify the expression
\[
\frac{5^{2x+4}}{25^{3x}}
\]
so that it has the form $b^{f(x)}$, where $b$ is a real number and $f(x)$ is some function of $x$. Using the properties of exponential functions, we obtain

\[
\frac{5^{2x+4}}{25^{3x}} = \frac{5^{2(x+2)}}{25^{3x}} = \frac{25^{x+2}}{25^{3x}} = \frac{25^{x+2}}{25^{3x}} = \frac{25^{x+2-3x}}{25^{2(1-x)}} = (25^2)^{1-x} = 625^{1-x}.
\]

It should be noted that we could have stopped at $25^{2-2x}$, since that is of the form $b^{f(x)}$, but in this case we continued so that we could have the simplest possible exponent. □

**Example 3.11** We will solve the equation

\[(x + 1)^2 = 16\]

for $x$. Since $a^x = b^y$ if and only if $a = b$, we raise both sides of this equation to the $1/2$ power, and obtain

\[[(x + 1)^2]^{1/2} = 16^{1/2},\]

which, by the properties of exponential functions, simplifies to

\[(x + 1)^1 = 16^{1/2},\]

or

\[x + 1 = 4.\]

It follows that the solution is $x = 3$. □

**Example 3.12** We will solve the equation

\[6^{2x^2} = 6^{2x-1}\]

for $x$. Since $a^x = a^y$ if and only if $x = y$ (provided that $a \neq 1$), it follows that

\[x^2 = 2x - 1.\]
Rearranging, we obtain the equation \( x^2 - 2x + 1 = 0 \), which can be factored to obtain \( (x - 1)^2 = 0 \). We conclude that the solution is \( x = 1 \). \( \square \)

**Example 3.13** We will solve the equation
\[
25^{3x} = 5^{x+1}
\]
for \( x \). Using the property that \( a^{xy} = (a^x)^y \), we can rewrite this equation as
\[
(5^2)^{3x} = 5^{x+1}
\]
or
\[
5^{6x} = 5^{x+1}.
\]
Since \( a^x = a^y \) if and only if \( x = y \), provided that \( a \neq 1 \), it follows that \( 6x = x + 1 \). Rearranging, we obtain the equation \( 5x = 1 \), which has the solution \( x = 1/5 \). \( \square \)

### 3.1.3 Applications

Exponential functions are useful in a number of practical applications, a few of which we will now discuss.

**Population Growth**

In any given population, it tends to be the case that factors that change the size of a population, particularly birth or death, tend to occur at a rate that is proportional to the size of the population. It follows that populations tend to grow exponentially, so a graph of population as a function of time looks somewhat like the graph of an exponential function. As we will see, this fact allows us to use current and past population data to estimate future population growth.

One way of modeling population growth using an exponential function is to use the *doubling time growth model* to measure the population, denoted by \( P \), at a given time \( t \). This model consists of the equation
\[
P = P_02^{t/d}, \tag{3.9}
\]
where \( P_0 \) is the population at \( t = 0 \), and \( d \) is the *doubling time*, which is the amount of time that is needed for the population to double. Any unit of time, such as hours or years, may be used for \( t \), but \( d \) must use the same unit as \( t \). If the doubling time and the initial population is known, then the population at any time \( t \) can be estimated using equation (3.9). Later in this book, we will learn how to solve more difficult problems involving population growth.
Radioactive Decay

A radioactive substance decays at a rate that is proportional to the amount that currently exists. It follows that radioactive decay over time, like population growth over time, can be modeled using an exponential function. The half-life decay model can be used to measure the amount \( A \) of a radioactive substance that will exist at time \( t \). The model consists of the equation

\[
A = A_0 \left( \frac{1}{2} \right)^{t/h},
\]

(3.10)

where \( A_0 \) is the amount of the substance that exists at \( t = 0 \), and \( h \) is the half-life, which is the amount of time that is needed for half of the substance to decay. If the half-life and the initial amount is known, then the amount at any time \( t \) can be measured using equation (3.10). Later in this book, we will learn how to solve more difficult problems involving radioactive decay.

Example 3.14 Suppose that a radioactive substance has a half-life of 50 years, and that initially, we have a 1 kg (1000 g) sample of the substance. How much of the substance, in grams, will remain after 150 years? We use the model for radioactive decay,

\[
A = A_0 \left( \frac{1}{2} \right)^{t/h},
\]

(3.11)

where \( A_0 \) is the initial amount, \( t \) is the amount of time that passes, \( A \) is the amount at time \( t \) (where \( t = 0 \) corresponds to the initial time), and \( h \) is the half-life. Setting \( A_0 = 1000 \), \( h = 50 \), and \( t = 150 \), we obtain

\[
A = 1,000 \left( \frac{1}{2} \right)^{150/50} = 1,000 \left( \frac{1}{2} \right)^3 = 1,000 \frac{1}{8} = 125.
\]

(3.12)

Compounding of Interest

When an amount of money, called the principal, is deposited into a savings account, or is loaned, it earns interest at a given interest rate, which is a percentage of the principal. The interest is added to the principal after a given period of time, at which point the combined amount of the principal and interest may be reinvested, and therefore earn additional interest. The interest that is paid on reinvested interest is called compound interest. Since
the interest is proportional to the principal, the growth of the combined amount of principal and compound interest over time can be modeled by an exponential function that has the form

\[ A = P \left(1 + \frac{r}{n}\right)^{nt}, \] (3.13)

where \( P \) is the principal that is originally invested, \( r \) is the annual interest rate, \( n \) is the number of times per year that interest is compounded (at regular intervals), \( t \) is the amount of time, measured in years, that the principal is invested, and \( A \) is the amount of money, including principal and compound interest, that exists after \( P \) has been invested for \( t \) years. Later in this chapter, we will learn how to use this model to determine how much interest can be earned when interest is compounded \textit{continuously}, instead of at regular intervals.

\textbf{Example 3.15} Suppose that $1,000 is deposited in a savings account that earns interest at the (unrealistically generous) rate of 10% per year. If interest is compounded annually, then, after one year, the balance in the account, denoted by \( A \), will be

\[ A = P(1 + r) = 1,000(1 + 0.1) = 1,100. \] (3.14)

After two years, the balance \( A \) will be

\[ A + P(1 + r)^2 = 1,000(1 + 0.1)^2 = 1,210, \] (3.15)

since the $100 interest that was earned during the first year also earns interest during the second year, resulting in the higher interest of $110 earned in the second year.

If interest is compounded twice a year, then, after one year, the balance \( A \) is

\[ A = P \left(1 + \frac{r}{2}\right)^2 = 1,000(1 + 0.05)^2 = 1,102.50, \] (3.16)

and if interest is compounded monthly (12 times a year), then the balance \( A \) after one year is

\[ A = P \left(1 + \frac{r}{12}\right)^{12} = 1,000 \left(1 + \frac{1}{120}\right)^{12} = 1,104.71. \] (3.17)

Finally, if interest is compounded continuously, then after one year, the balance \( A \) is

\[ A = Pe^{rt} = 1,000e^{0.1} = 1000e^{0.1} = 1,105.17. \] (3.18)
### Example 3.16

In this example, we derive the formula for the total balance after the principal, $P$, has been earning interest at a rate $r$ for one year, with interest compounded every six months. After six months, interest is compounded for the first time. Therefore, the balance $A$ is

$$A = P + \frac{r}{2}P = P \left(1 + \frac{r}{2}\right), \quad (3.19)$$

where the factor of $1/2$ appears because only half a year has elapsed.

After a full year has passed, interest is compounded a second time. Since six months have passed since the previous compounding, the interest is computed by multiplying the balance after six months by $r/2$, just as the original balance was multiplied by $r/2$ to compute the interest earned during the first six months. The balance $A$ after one year is

$$A = P \left(1 + \frac{r}{2}\right) + \frac{r}{2}P \left(1 + \frac{r}{2}\right)$$

$$= P \left[ \left(1 + \frac{r}{2}\right) + \frac{r}{2} \left(1 + \frac{r}{2}\right) \right]$$

$$= P \left(1 + \frac{r}{2}\right) \left(1 + \frac{r}{2}\right)$$

$$= P \left(1 + \frac{r}{2}\right)^2. \quad (3.20)$$

In the second step, $P$ is factored out, and in the third step, $1 + r/2$ is factored out.

It should be noted that this same formula can be obtained from the general formula for compounding of interest,

$$A = P \left(1 + \frac{r}{n}\right)^{nt}, \quad (3.21)$$

by setting $t = 1$, since we are computing the balance after one year has passed, and $n = 2$, since we are compounding twice per year. □

### 3.2 The Natural Exponential Function

In the previous section, we learned about the exponential function $f(x) = b^x$, where the base $b$ was an arbitrary positive number (excluding $b = 1$). In this section, we focus on a specific choice of $b$ that has proven to be particularly interesting and useful.
3.2.1 Continuously Compounded Interest

Recall from the previous section that when an amount of money $P$, called the principal, earns interest at an annual interest rate $r$, then the value of the investment after $t$ years is $A$, where

$$A = P \left(1 + \frac{r}{n}\right)^{nt}, \quad (3.22)$$

where $n$ is the number of times per year that interest is compounded. In this section, we consider the following question: how is the value of the investment influenced by the number $n$? To answer this question, we use the laws of exponents to rewrite equation (3.22) as follows:

$$A = P \left(1 + \frac{r}{n}\right)^{n/r}t = P \left[\left(1 + \frac{r}{n}\right)^{n/r}\right]^{rt}. \quad (3.23)$$

If we make the substitution $m = n/r$, then this simplifies to

$$A = P \left[\left(1 + \frac{1}{m}\right)^{m}\right]^{rt}. \quad (3.24)$$

As interest is compounded more often per year, the number $n$ increases, which implies that $m$ increases as well. What happens if $m$ becomes arbitrarily large? It turns out that the expression $(1+1/m)^m$ actually converges to a number $e$ whose value is approximately

$$e = 2.718281828459... \quad (3.25)$$

where the ... indicates that the decimal expansion continues indefinitely. This number $e$ is actually an irrational number, like $\pi$, meaning that it cannot be written as a fraction.

We conclude that as interest is compounded more often, the value $A$ of the investment after $t$ years approaches the limit

$$A = Pe^{rt}. \quad (3.26)$$

It should be noted that as $n$ increases, the quantity $(1+1/m)^m$ increases as well, with the number $e$ being its upper limit. It follows that the value of the investment increases more rapidly if interest is compounded more often, with continuous compounding providing the fastest possible growth.

We see the amount by which the investment grows over time is determined by the exponential function with base $e$. We now state the definition of this function precisely.
3.2. THE NATURAL EXPONENTIAL FUNCTION

Definition 3.2 The exponential function with base $e$, or the natural exponential function, is the function $f(x)$ defined by

$$f(x) = e^x,$$

(3.27)

where $x$ is any real number and $e$ is the number that is approached by $(1 + 1/m)^m$, as $m$ becomes arbitrarily large.

3.2.2 Rate of Change of the Exponential Function with Base $e$

As stated in the preceding definition, the exponential function with base $e$ is also known as the “natural” exponential function. This characterization of the base $e$ as a natural base is connected to the rate at which $e^x$ changes as $x$ changes.

Recall that the rate of change of a linear function $f(x) = mx + b$, where $m$ and $b$ are constants, is equal to the slope of the line $y = mx + b$. The slope is the ratio of the change in $y$ to the change in $x$, and is equal to the constant $m$. The rate of change of any other type of function can be defined in a similar fashion, in terms of the ratio of the change in $f(x)$ to the change in $x$, except that the rate of change varies as $x$ varies.

Any exponential function of the form $f(x) = b^x$ has the property that the rate of change of $b^x$ with respect to $x$ is proportional to $b^x$. Informally, this implies that for any real number $x_0$,

$$\frac{b^{x_0+\Delta x} - b^{x_0}}{\Delta x} \approx cb^{x_0}$$

(3.28)

where $\Delta x$ is an “infinitesimally small” nonzero number and $c$ is a constant. Later in this chapter, we will be able to state the value of $c$, but for now, we state that when $b = e$, the constant $c$ is equal to 1. That is, the rate of change of $e^x$, with respect to $x$, is equal to $e^x$. This equality between the function and its rate of change is illustrated in Figure 3.3.

3.2.3 Exponential Growth Phenomena

The exponential function with base $e$ is used to model a wide variety of phenomena. We now list the most common models that involve this function.

- **Exponential Growth:** If a quantity $y = f(t)$ is known to increase at a rate that is equal to $ky$ as $t$ changes, where $k$ is a positive constant, then $y$ has the form $y = ce^{kt}$, where $c$ is the value of $y$ when $t = 0$. 

Figure 3.3: The rate of change of $e^x$, the exponential function with base $e$, with respect to $x$ is equal to $e^x$. In the left plot, observe that $e^0 = 1$, and the slope of the line that is parallel to the graph of $e^x$ at $x = 0$ is also equal to 1. In the right plot, observe that $e^1 = e$, and the slope of the line that is parallel to the graph of $e^x$ at $x = 1$ is also equal to $e$.

This model is useful for studying short-term population growth or continuously compounded interest.

- **Exponential Decay**: If a quantity $y = f(t)$ is known to decrease at a rate that is equal to $ky$ as $t$ changes, where $k$ is a positive constant, then $y$ has the form $y = ce^{-kt}$, where $c$ is the value of $y$ when $t = 0$. This model is useful for studying radioactive decay and light absorption.

- **Limited Growth**: If a quantity $y = f(t)$ is known to increase at a rate that is equal to $k(c - y)$ as $t$ changes, where $k$ and $c$ are positive constants, then $y$ has the form $y = c(1 - e^{-kt})$. A function of this form grows at a slower and slower rate as $y$ approaches the value $c$, and $y$
never surpasses $c$. This model is useful for studying learning skills or sales fads.

- **Logistic Growth:** If a quantity $y = f(t)$ is known to increase at a rate that is proportional to $y(M - y)$ as $t$ changes, where $M$ is a positive constant, then $y$ has the form $y = M(1 + ce^{-kt})$, where $c$ and $k$ are positive constants. A function of this form grows exponentially until it approaches the value $M$, at which point the growth slows, and $y$ never surpasses $M$. This model is useful for studying long-term population growth, epidemics, or sales of new products.

Examples of all of these models are shown in Figure 3.4.

## 3.3 Inverse Functions

There are an ever-increasing number of applications in which it is necessary to solve problems that are known as inverse problems. In such problems, the traditional roles of input and output are reversed. Instead of asking what is the output from a given input, one may instead know the output already, and the input needs to be determined. The original problem is referred to as a forward problem, while the new formulation is called an inverse problem.

We are already aware of some inverse problems. If one considers the problem of multiplying two numbers the forward problem, then dividing is the inverse problem. If differentiating a function is the forward problem, then anti-differentiating is the inverse problem. In many scientific and engineering applications, the forward problem consists of determining the response to some stimulus. The corresponding inverse problem is: given a response that has been measured, what is the stimulus that produced the response?

In many cases, the solution to some forward problem may be a function $y = f(x)$, where $x$ is the input and $y$ is the output. The corresponding inverse problem consists of determining $x$ from any given $y$. The solution of this problem, if it exists, is called the inverse function of $f$, and is denoted $f^{-1}$ (not to be confused with the reciprocal of $f$). It is defined by the relation

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

When does the inverse function exist? Recall that a function $f(x)$ is defined to be a set of ordered pairs of the form $(x, y)$ where $x$ is in the domain of $f$ and $y$ is in the range of $f$. For $f$ to be a function, it must have the property that for every $x$ in the domain, there is exactly one element...
Figure 3.4: Examples of various models of exponential growth and decay. All constants in the following descriptions are assumed to be positive. Top left: exponential growth is modeled by functions of the form $ce^{kt}$. Top right: exponential decay is modeled by functions of the form $ce^{-kt}$. Bottom left: limited growth is modeled by functions of the form $c(1 - e^{-kt})$. Bottom right: logistic growth is modeled by functions of the form $M/(1 + ce^{-kt})$.

For the inverse function to exist, the opposite must be true: for each element $y$ in the range of $f$, which is the domain of $f^{-1}$, there must be exactly one element $x$ in the domain of $f$, which is the range of $f^{-1}$, such that $x = f^{-1}(y)$.

Now, suppose that a function $f(x)$ has the property that for two different elements $x_1$ and $x_2$ in its domain, $f(x_1) = f(x_2)$. Then, $f$ cannot have an inverse function, since the value $y = f(x_1)$ in the domain of $f^{-1}$ corresponds to two values in the range of $f^{-1}$, $x_1$ and $x_2$, and therefore $f^{-1}$ cannot be a function.

It follows that in order for $f$ to have an inverse function, it must be
3.3. INVERSE FUNCTIONS

One-to-one: if \( x_1 \neq x_2 \), then \( f(x_1) \neq f(x_2) \). If this is the case, then the inverse function exists. There is a simple test that can be used to determine visually if \( f \) is one-to-one, known as the horizontal line test. If it is possible to draw a horizontal line that intersects the graph of \( f \) more than once, then \( f \) cannot be one-to-one, since the horizontal line corresponds to an element in the range of \( f \) that is associated with more than one element in the domain.

Example 3.17 The following function \( f(x) \) is not one-to-one, since \( f(2) = f(6) = 2 \).

\[
\begin{array}{cccccc}
 x & 1 & 2 & 3 & 4 & 5 & 6 \\
f(x) & 1.5 & 2 & 3.6 & 5.3 & 2.8 & 2 \\
\end{array}
\]

\[\square\]

Example 3.18 The following function \( f(x) \) is one-to-one.

\[
\begin{array}{cccccc}
 x & 1 & 2 & 3 & 4 & 5 & 6 \\
f(x) & 1 & 2 & 4 & 8 & 16 & 32 \\
\end{array}
\]

\[\square\]

Example 3.19 Let \( f(t) \) describe the height of a kicked football at time \( t \). This function is not one-to-one, since the football must eventually come back down after gaining altitude immediately after being kicked. \[\square\]

The following result provides an easy way of determining whether a function is one-to-one, and therefore would have an inverse. Recall that a function \( f \) is increasing if \( f(x) > f(y) \) whenever \( x > y \), and decreasing if \( f(x) < f(y) \) whenever \( x > y \).

Theorem 3.1 If a function \( f \) is increasing on its entire domain, or decreasing on its entire domain, then \( f \) is one-to-one.

Proof Suppose that \( f \) is increasing. Then, if \( x < y \), then \( f(x) < f(y) \). Let \( x \) and \( y \) be any two real numbers in the domain of \( f \) such that \( x \neq y \). Then, we must have \( x < y \) or \( x > y \). If \( x < y \), then \( f(x) < f(y) \), and if \( x > y \), then \( f(x) > f(y) \). Therefore, \( f(x) \neq f(y) \). Since this is true for any \( x \) and \( y \) in the domain of \( f \), we conclude that \( f \) is one-to-one. The proof for the case where \( f \) is decreasing is similar. \[\square\]
If $f$ has an inverse function, how can it be determined? This is not always easy to do, but the general procedure is to solve the equation $y = f(x)$ for $x$, thus obtaining the relation $x = f^{-1}(y)$. The graph of $f^{-1}$, however, is easy to obtain from the graph of $f$. If the graph of $f$ is reflected about the line $y = x$, then, in the new graph, the roles of $x$ and $y$ are reversed, and the curve that is the reflection of the graph of $f$ is now the graph of $f^{-1}$. If the inverse function is difficult to determine algebraically, this graph may be helpful to, at least, approximate $f^{-1}$.

**Example 3.20** Let $f(x) = 3x + 4$. Since $f$ is a linear function, with nonzero slope, it is one-to-one, and therefore has an inverse function $f^{-1}$. To find $f^{-1}$, we can use the property that $y = f(x)$ if and only if $x = f^{-1}(y)$. We begin with the equation $y = f(x)$, which, in this case, is the equation

$$y = 3x + 4. \quad (3.29)$$

We then solve this equation for $x$, which yields

$$x = \frac{y - 4}{3}. \quad (3.30)$$

Since $y = f(x)$ implies that $x = f^{-1}(y)$, we conclude that the inverse function is given by

$$f^{-1}(y) = \frac{y - 4}{3}. \quad (3.31)$$

\[\Box\]

**Example 3.21** Let $f(x) = x^2$. This function is not one-to-one if we allow $x$ to be any real number, since $f(-x) = (-x)^2 = x^2 = f(x)$. However, if we restrict the domain of $f$ to be the interval $[0, \infty)$ (that is, the set of all $x$ so that $x \geq 0$), then $f$ is one-to-one, and therefore has an inverse function $f^{-1}$.

To find $f^{-1}$, we proceed as in the previous example and start with the equation $y = x^2$, where we assume that $x \geq 0$. We can solve this equation for $x$ by raising both sides to the $1/2$ power, which yields

$$y^{1/2} = (x^2)^{1/2} = x^{2(1/2)} = x^1 = x. \quad (3.32)$$

We conclude that $f^{-1}$ is described by $f^{-1}(y) = y^{1/2} = +\sqrt{y}$.

The graphs of $f$ and $f^{-1}$ are shown in Figure 3.5. Note that the graph of $f^{-1}$ can be obtained from the graph of $f$ by reflecting the graph of $f$ across the diagonal line $y = x$. Also note that both $f$ and $f^{-1}$ are increasing functions. \[\Box\]
3.3. INVERSE FUNCTIONS

Figure 3.5: Graph of \( y = x^2 \) (solid curve) and its inverse function, \( y = +\sqrt{x} \) (dashed curve)

Example 3.22 Let \( f(x) = 1 + 2x + 3x^2 \), on the domain \(-1/3 < x < \infty\). Show that \( f(x) \) is one-to-one on this domain and the inverse function.

Solution We have \( f'(x) = 2 + 6x \). It follows that for \( x > 1/3 \), \( f'(x) > 0 \), and therefore \( f(x) \) is increasing for all \( x \) in the given domain. This means that it is not possible for \( f(x) \) to assume a value more than once on this domain (though \( f(x) \) is not one-to-one on the domain \(-\infty < x < \infty\), since it is a parabola).

To find its inverse, we solve the equation

\[
y = 1 + 2x + 3x^2
\]

for \( x \). We use the quadratic formula to solve the equation

\[
3x^2 + 2x + (1 - y) = 0
\]
and obtain
\[ x = \frac{-2 \pm \sqrt{2^2 - 4(3)(1 - y)}}{2(3)} \]
which simplifies to
\[ x = -\frac{1}{3} \pm \frac{\sqrt{3y - 2}}{3}. \]
Since the domain of \( f(x) \) is the interval \((-1/3, \infty)\), it follows that the range of the inverse function defined by \( x = f^{-1}(y) \) must also be this interval. Therefore, we must choose the positive square root, and we have
\[ f^{-1}(y) = -\frac{1}{3} + \frac{\sqrt{3y - 2}}{3}, \]
which, if desired, can also be written as
\[ f^{-1}(x) = -\frac{1}{3} + \frac{\sqrt{3x - 2}}{3}. \]

\[ \square \]

Even if a formula for the inverse function is not available, it is still possible to determine its derivative at a given point. Intuitively, if \( b = f(a) \) and \( f'(a) = m \), then the slope of the tangent line of \( f \) at \( x = a \) has slope \( m \), and therefore the slope of the tangent line of \( f^{-1} \) at \( y = b \) should have slope \( 1/m \); in other words, \((f^{-1})'(b) = 1/m\). This should be true because reflecting the tangent line about the line \( y = x \) has the effect of taking the reciprocal of the slope; i.e., the roles of “rise” and “run” are reversed, where slope is, conceptually, “rise over run”.

To determine if this is the case in general, we can rely on the cancellation equations, which state that
\[ f(f^{-1}(y)) = y \quad \text{and} \quad f^{-1}(f(x)) = x \]
for every \( y \) in the range of \( f \) and every \( x \) in the domain of \( f \). Taking the first cancellation equation and differentiating both sides with respect to \( y \) via the chain rule, we obtain
\[ f'(f^{-1}(y)) \cdot (f^{-1})'(y) = 1 \]
which yields
\[ (f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}. \]
3.4. **GENERAL LOGARITHMIC FUNCTIONS**

In other words, the derivative of the inverse function is, in a sense, the reciprocal of the derivative. However, it should be noted that \( f' \) must be evaluated at \( f^{-1}(y) \), not \( y \).

It is important to note that the derivative of \( f^{-1} \) is only defined at \( y \) if \( f'(f^{-1}(y)) \neq 0 \); in other words, \( f \) must not have a horizontal tangent at \( f^{-1}(y) \). This only makes sense because this horizontal tangent of \( f \) would, by reflection, correspond to a vertical tangent of \( f^{-1} \) at \( y \), meaning that \( f^{-1} \) should not be differentiable at \( y \) anyway.

**Example 3.23** Let \( f(x) = x^2 + 1 \), on the domain \([0, \infty)\). Compute \((f^{-1})'(2)\) without computing \( f^{-1} \) or using the differentiation rule for inverse functions.

**Solution** Setting \( x = f^{-1}(2) \) and applying the cancellation equations yields \( f(x) = x^2 + 1 = 2 \), which has the solution \( x = \pm 1 \). Since the domain of \( f(x) \) for this problem is \([0, \infty)\), we must have \( f^{-1}(2) = 1 \). At \( x = 1 \), the slope of the tangent line to the graph of \( f \) is equal to 2, since \( f'(x) = 2x \) which yields \( f'(1) = 2 \). The graph of \( f^{-1} \) can be obtained by switching the \( x \) and \( y \) coordinates of points on the graph of \( f \), so this line that is tangent to the graph of \( f \) at the point \((1, 2)\) corresponds to a line that is tangent to the graph of \( f^{-1} \) at the point \((2, 1)\). The tangent line for \( f^{-1} \) therefore must have slope \( 1/2 \), since slope is defined to be the ratio of change in \( y \) to change in \( x \), and the roles of \( x \) and \( y \) are reversed in graphing \( f^{-1} \) or any of its tangent lines. \( \square \)

### 3.4 General Logarithmic Functions

In this section, we discuss the inverse function of the exponential function with base \( b \), \( f(x) = b^x \). As discussed in Section 3.1, \( b^x \) is an increasing function if \( b > 1 \), and it is a decreasing function if \( 0 < b < 1 \). In either case, by Theorem 3.1, \( b^x \) is one-to-one. It follows that \( b^x \) has an inverse function. We now precisely define this inverse function.

**Definition 3.3** *(Logarithmic Function)* Let \( b > 0 \), and assume that \( b \neq 1 \). Let \( f(y) = b^y \) be the exponential function with base \( b \). The **logarithmic function with base \( b \)** is the inverse function of \( f \), and is denoted by \( f^{-1}(x) = \log_b x \). That is,

\[
y = \log_b x \quad \text{if and only if} \quad b^y = x, \quad (3.33)
\]

where \( x > 0 \) and \( y \) is a real number. The statement \( y = \log_b x \) in equation (3.33) is called the **logarithmic form**, and the statement \( b^y = x \) in equation
(3.33) is called the exponential form. The number $y$ in equation (3.33) is called the logarithm, or log, to base $b$ of $x$.

It is important to realize that the value of $\log_b x$ is an exponent. Specifically, if $y = \log_b x$, then $y$ is the exponent to which $b$ must be raised in order to obtain $x$.

**Example 3.24** Because the functions $f(x) = 2^x$ and $f^{-1}(x) = \log_2 x$ are inverses of one another, it follows that

$$2^{\log_2 64} = 64,$$

and

$$\log_2 2^6 = 6.$$ \hspace{1cm} (3.34) \hspace{1cm} (3.35)

Both of these equations can be verified by noting that $2^6 = 64$. $\square$

The definition provides a very useful statement for solving equations involving exponential and logarithmic functions: the equivalence of the exponential form $b^y = x$ and the logarithmic form $y = \log_b x$. It is important to be able to convert between these forms in order to solve such equations.

**Example 3.25** The equation $4^3 = 64$ is in exponential form. The corresponding logarithmic form of this equation is $\log_4 64 = 3$. $\square$

**Example 3.26** The equation $10^{-3} = 0.001$ is in exponential form. The corresponding logarithmic form of this equation is $\log_{10} 0.001 = -3$. $\square$

### 3.4.1 Properties of Logarithmic Functions

The properties of exponential functions introduced in Section 3.1 can be used to derive similar properties for logarithmic functions. We will see that these properties are extremely useful in solving problems involving logarithmic functions.

In the statement of the following theorem that lists these properties, we use the symbol $\iff$, which means “if and only if.” That is, the statement $p \iff q$ means that the statements $p$ and $q$ are true under exactly the same circumstances, so, in effect, they are equivalent statements.

**Theorem 3.2** (Properties of Logarithmic Functions) Let $b$, $x$, $y$, and $r$ be real numbers, where $b$, $x$ and $y$ are positive, and $b \neq 1$. Then
3.4. **GENERAL LOGARITHMIC FUNCTIONS**

1. \( \log_b 1 = 0 \)

2. \( \log_b b = 1 \)

3. \( \log_b b^r = r \)

4. \( b^{\log_b r} = r \), where \( r > 0 \).

5. \( \log_b xy = \log_b x + \log_b y \)

6. \( \log_b \frac{x}{y} = \log_b x - \log_b y \)

7. \( \log_b x^r = r \log_b x \)

8. \( \log_b x = \log_b y \iff x = y \)

**Proof**

1. Since \( b > 0 \), \( b^0 = 1 \). Rewriting this equation in logarithmic form immediately yields \( \log_b 1 = 0 \).

2. Since \( b > 0 \), \( b^1 = b \). Rewriting this equation in logarithmic form immediately yields \( \log_b b = 1 \).

3. Let \( z = b^r \). If we rewrite this equation in logarithmic form, we obtain \( \log_b z = r \). Substituting \( b^r \) for \( z \) yields \( \log_b b^r = r \).

   Alternatively, let \( f(r) = b^r \). Then the inverse function of \( f \) is \( f^{-1}(z) = \log_b z \). Because \( f \) and \( f^{-1} \) are inverse functions of one another, \( f^{-1}(f(r)) = r \). However, by the definition of these functions, we also have \( f^{-1}(f(r)) = f^{-1}(b^r) = \log_b b^r \). Therefore \( \log_b b^r = r \).
4. Let \( z = \log_b r \). Note that the definition of \( z \) makes sense only if \( r > 0 \), which we assume in this case. Then, rewriting the equation \( z = \log_b r \) in exponential form yields \( b^z = r \). Substituting \( z = \log_b r \) into the exponential form yields \( b^{\log_b r} = r \).

Alternatively, let \( f(z) = b^z \). Then the inverse function of \( f \) is \( f^{-1}(r) = \log_b r \). Because \( f \) and \( f^{-1} \) are inverse functions of one another, \( f(f^{-1}(r)) = r \). However, by the definition of these functions, we also have \( f(f^{-1}(r)) = f(\log_b r) = b^{\log_b r} \). Therefore \( \log_b b^r = r \).

5. Let \( s = \log_b x \) and \( t = \log_b y \). Then, rewriting the definitions of \( s \) and \( t \) in exponential form, we obtain \( x = b^s \) and \( y = b^t \). Recall that for any real numbers \( s \) and \( t \), \( b^{s+t} = b^s b^t \). This implies that \( b^{s+t} = xy \). If we rewrite this equation in logarithmic form, we obtain \( \log_b xy = s + t \). By the definitions of \( s \) and \( t \), we obtain \( \log_b xy = \log_b x + \log_b y \).

6. Let \( s = \log_b x \) and \( t = \log_b y \). Then, rewriting the definitions of \( s \) and \( t \) in exponential form, we obtain \( x = b^s \) and \( y = b^t \). Recall that for any real numbers \( s \) and \( t \), \( b^{s-t} = b^s / b^t \). This implies that \( b^{s-t} = x/y \). If we rewrite this equation in logarithmic form, we obtain \( \log_b x/y = s - t \). By the definitions of \( s \) and \( t \), we obtain \( \log_b x/y = \log_b x - \log_b y \).

7. Let \( s = \log_b x \). Then, rewriting the definition of \( s \) in exponential form, we obtain \( x = b^s \). Recall that for any real numbers \( r \) and \( s \), \( b^{sr} = (b^s)^{r} \). This implies that \( b^{sr} = x^r \). If we rewrite this equation in logarithmic form, we obtain \( \log_b x^r = sr \). By the definitions of \( s \) and \( t \), we obtain \( \log_b x^r = r \log_b x \).

8. Let \( s = \log_b x \) and \( t = \log_b y \). Then, rewriting the definitions of \( s \) and \( t \) in exponential form, we obtain \( x = b^s \) and \( y = b^t \). Recall that for any real numbers \( s \) and \( t \), \( b^s = b^t \) if and only if \( s = t \). This implies that \( x = y \) if and only if \( s = t \). By the definitions of \( s \) and \( t \), we conclude that \( x = y \) if and only if \( \log_b x = \log_b y \). This is logically equivalent to the statement that \( \log_b x = \log_b y \) if and only if \( x = y \).

It is important to use caution when applying these properties. The following mistakes are very common:

- **Confusing the log of a sum or difference with the sum or difference of logs.** The expression \( \log_b x + \log_b y \) is a sum of logs, and it can be simplified to \( \log_b xy \). The expression \( \log_b(x + y) \) is a log of a sum, and it cannot be simplified.
3.5. THE NATURAL AND COMMON LOGARITHMIC FUNCTIONS

- Confusing the log of a product or quotient with the product or quotient of logs. The expression \((\log_b x)^n\), where \(n\) is a positive integer, is a product of logs, and it cannot be simplified. The expression \(\log_b x^n\) is a log of a product, and it can be simplified to \(n \log_b x\). Similarly, the expression \(\log_b x/\log_b y\) is a quotient of logs, and it cannot be simplified. On the other hand, the expression \(\log_b (x/y)\) is a log of a quotient, and it can be rewritten as \(\log_b x - \log_b y\).

3.4.2 Graphs of Logarithmic Functions

The graph of \(\log_b x\) can be obtained from the graph of \(b^x\) by interchanging the \(x\)- and \(y\)-coordinates of each point on the graph of \(b^x\), since \(\log_b x\) is its inverse function. It follows that if \(b > 1\), then \(\log_b x\) is an increasing function, and if \(0 < b < 1\), then \(\log_b x\) is a decreasing function. Furthermore, the graph of every logarithmic function passes through the point \((1, 0)\), just as the graph of every exponential function passes through the point \((0, 1)\).

Finally, the graph of \(\log_b x\) approaches the vertical line \(x = 0\), just as the graph of \(b^x\) approaches the horizontal line \(y = 0\). In terms of limits, we have

\[
\lim_{x \to \infty} \log_b x = \infty, \quad \lim_{x \to 0^+} \log_b x = -\infty
\]

if \(b > 1\), and

\[
\lim_{x \to \infty} \log_b x = -\infty, \quad \lim_{x \to 0^+} \log_b x = \infty
\]

if \(b < 1\).

3.5 The Natural and Common Logarithmic Functions

In the previous section, we defined the logarithmic function with base \(b\), where \(b\) is any positive real number other than one. In this section, we focus on the two most frequently used bases.

3.5.1 Definition and Evaluation

The most frequently used base for exponentiation, across all applications areas in which mathematics plays a significant role, is certainly base 10. The logarithmic function of base 10 is often useful for solving equations involving exponentiation with base 10. Therefore, this logarithmic function, \(\log_{10} x\), is called the common logarithm, and is often written as simply \(\log x\).
In Section 3.2, we introduced the natural exponential function, which was of base \( e \). Because of this function’s importance, we similarly define the natural logarithmic function, or natural logarithm, to be the logarithmic function with base \( e \). This function, \( \log_e x \), is written as \( \ln x \), where the “n” in the symbol \( \ln \) means “natural.”

**Example 3.27** Simplify the expression

\[
\log_3 81
\]

(3.36)
to a single number.

**Solution** We use the fact that \( 3^4 = 81 \) and rewrite the expression as

\[
\log_3 3^4.
\]

(3.37)
Using the property \( \log_b x^y = y \log_b x \), we obtain \( 4 \log_3 3 \). Because \( \log_b b = 1 \) for any base \( b \), it follows that the value of the expression is simply 4.

It should be noted that the equation \( \log_3 81 = 4 \) is the logarithmic form of the equation \( 3^4 = 81 \), which is in exponential form. □

**Example 3.28** Simplify the expression

\[
\ln 6 + \ln 8 + 2 \ln 5 - 4 \ln 2
\]

(3.38)
so that it has the form \( \ln x \), for some number \( x \).

**Solution** Using the property \( \ln x^y = y \ln x \), we obtain

\[
\ln 6 + \ln 8 + \ln 25 - \ln 16.
\]

(3.39)
Using the property \( \ln xy = \ln x + \ln y \) yields

\[
\ln 48 + \ln 25 - \ln 16.
\]

(3.40)
Using this property again, along with the property \( \ln(x/y) = \ln x - \ln y \), we obtain

\[
\ln \frac{48(25)}{16}.
\]

(3.41)
However, \( 48 = 16 \cdot 3 \), so we have

\[
\ln \frac{48(25)}{16} = \ln 3 \cdot 25 = \ln 75.
\]

(3.42)
Example 3.29 Write the expression
\[ \ln \frac{x^2(x + 1)}{\sqrt{x^3 - 1}} \] (3.43)
in terms of natural logarithms of simpler functions.

Solution Using the property \( \ln(x/y) = \ln x - \ln y \) yields
\[ \ln x^2(x + 1) - \ln \sqrt{x^3 - 1}. \] (3.44)
Using the property \( \ln xy = \ln x + \ln y \) yields
\[ \ln x^2 + \ln(x + 1) - \ln \sqrt{x^3 - 1}. \] (3.45)
Using the property \( \ln x^y = y \ln x \) yields
\[ 2 \ln x + \ln(x + 1) - \frac{1}{2} \ln(x^3 - 1), \] (3.46)
since \( \sqrt{x} = x^{1/2} \).

In general, the polynomial \( x^n - 1 \) can be factored as follows:
\[ x^n - 1 = (x - 1)(x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1). \] (3.47)
It follows that \( x^3 - 1 = (x - 1)(x^2 + x + 1) \), and our expression becomes
\[ 2 \ln x + \ln(x + 1) - \frac{1}{2} \ln(x - 1)(x^2 + x + 1), \] (3.48)
which can be rewritten as
\[ 2 \ln x + \ln(x + 1) - \frac{1}{2} \ln(x - 1) - \frac{1}{2} \ln(x^2 + x + 1). \] (3.49)

Example 3.30 Solve the equation
\[ \log x = \frac{2}{3} \] (3.50)
for \( x \).

Solution This equation is in logarithmic form. Rewriting it in exponential form yields
\[ 10^{2/3} = x. \] (3.51)
Alternatively, we can obtain this answer from the original equation by applying the function \( 10^x \) to both sides. Since \( 10^x \) is the inverse function of \( \log x \), it follows that \( 10^{\log x} = x \). This is an example of the cancellation equation \( f(f^{-1}(x)) = x \). \( \square \)
Example 3.31 Solve the equation
\[ \log(x + 1) + \log(x - 1) = \log 3 \] (3.52)
for \( x \).

Solution Using the property \( \log xy = \log x + \log y \), we obtain
\[ \log(x + 1)(x - 1) = \log 3, \] (3.53)
which simplifies to
\[ \log(x^2 - 1) = \log 3. \] (3.54)
Since \( \log x = \log y \) if and only if \( x = y \), it follows that \( x^2 - 1 = 3 \), or \( x^2 = 4 \). This equation yields the solutions \( x = \pm 2 \).

However, the solution \( x = -2 \) is not a solution of the original equation (3.52). If we substitute \( x = -2 \) into this equation, we obtain the expressions \( \log(-1) \) and \( \log(-3) \), but these make no sense because logarithmic functions are only defined for positive real numbers. We conclude that the only solution of this equation is \( x = 2 \). \( \square \)

3.5.2 The Change-of-Base Formula

The natural logarithmic function \( \ln x \) provides a simple way to define logarithmic functions of other bases. From the identity
\[ x = b^{\log_b x}, \]
we take the natural logarithm of both sides to obtain
\[ \ln x = \ln b^{\log_b x}. \]
From the property \( \log_b x^y = y \log_b x \), we obtain
\[ \ln x = \log_b x \ln b. \]
Finally, we obtain the change-of-base formula
\[ \log_b x = \frac{\ln x}{\ln b}. \]
It can be seen that there is a simple relationship among logarithmic functions of all bases—they are all constant multiples of one another.
3.5.3 Applications

We now discuss some applications of the common logarithm, \( \log x \). In general, the common logarithm is particularly useful for measuring quantities that tend to vary over an extremely wide range of values, spanning several orders of magnitude, such as the range of real numbers from \( 10^{-10} \) to \( 10^{10} \). For such a quantity, the common logarithm provides a reasonably-sized number that indicates the magnitude of the quantity’s value, because it represents the value’s exponent, the power to which 10 must be raised to obtain this value.

It should also be noted that the common logarithm, because its base is 10, has a useful interpretation in terms of the representation of a real number as a sequence of decimal digits and a decimal point. Suppose that \( y = \log x \). If \( x > 1 \), then \( x \) has at least \( \lceil y \rceil \) significant digits to the left of the decimal point, where \( \lceil y \rceil \) is the smallest integer that is greater than or equal to \( y \). On the other hand, if \( x < 1 \), then \( y < 0 \), and \( x \) has \( \lceil -y \rceil - 1 \) zeros occurring to the right of the decimal point before the first significant digit.

**Example 3.32** If \( x = 20 \), then \( \log x = 1.301 \). Note that \( x \) has two digits to the left of the decimal point, and \( \lceil 1.301 \rceil = 2 \). On the other hand, if \( x = 0.002 \), then \( \log x = -2.699 \). Note that \( x \) has two zeros to the right of the decimal point before the first significant digit, and \( \lceil -(2.699) \rceil - 1 = 3 - 1 = 2 \).

**Sound Intensity**

The intensity of sound is measured in watts per square meter. With this choice of unit of measurement, the intensity of sounds that can be heard by the human ear without causing damage to the eardrum ranges from \( 10^{-12} \), for the softest audible sound, to \( 10^{0} \), the threshold for pain. Because of this wide range, the intensity is converted to an alternative measurement, a decibel level. The decibel level of a sound, denoted by \( D \), is defined by

\[
D = 10 \log \frac{I}{I_0},
\]

where \( I \) is the intensity of the sound, and \( I_0 = 10^{-12} \), the intensity of the least audible sound to a healthy human ear. It follows that this least audible sound has a decibel level of zero, while a sound of intensity \( 10^{0} \), the threshold for pain, has a decibel level of

\[
D = 10 \log \frac{10^{0}}{10^{-12}} = 10 \log 10^{12} = 10 \cdot 12 \log 10 = 10 \cdot 12 = 120. \quad (3.56)
\]
Certainly, the range of numbers from 0 to 120 is much easier to work with than the range of numbers between $10^{-12}$ and $10^0$.

**Earthquake Intensity**

Earthquake intensity, measured in joules, also assumes a very wide range of values. Therefore, like sound intensity, earthquake intensity is converted to *magnitude on the Richter scale*. This magnitude, denoted by $M$, is defined by

$$M = \frac{2}{3} \log \frac{E}{E_0},$$

where $E$ is the earthquake intensity, measured in joules, and $E_0$ is the intensity of a very small “reference earthquake”, $10^{4.40}$ joules. It follows from this definition of magnitude that an earthquake of magnitude greater than 7.5 on the Richter scale is quite severe. In fact, it is 1000 times more intense than an earthquake of magnitude 5.5.

**Example 3.33** Suppose that a sound has intensity $I = 2 \times 10^{-6}$ watts per square meter. What is the decibel level of the sound?

**Solution** We use the formula

$$D = 10 \log \frac{I}{I_0},$$

where $I$ is the intensity, $D$ is the decibel level, and $I_0 = 10^{-12}$ watts per square meter, which is the intensity of the softest audible sound. We substitute $I = 2 \times 10^{-6}$ into equation (3.58) and use the properties of logarithms to obtain

$$D = 10 \log \frac{2 \times 10^{-6}}{10^{-12}}$$

$$= 10[\log 2 \times 10^{-6} - \log 10^{-12}]$$

$$= 10[\log 2 + \log 10^{-6} - \log 10^{-12}]$$

$$= 10[\log 2 + (-6) \log 10 - (-12) \log 10]$$

$$= 10[\log 2 - 6 + 12]$$

$$= 10[6 + \log 2]$$

$$\approx 10[6 + 0.301]$$

$$\approx 63.01.$$  

(3.59)
Example 3.34 The 1906 San Francisco earthquake had an intensity of approximately $E = 6 \times 10^{16}$ joules. What was the magnitude of this quake on the Richter scale?

Solution We use the formula

$$M = \frac{2}{3} \log \frac{E}{E_0},$$

where $E$ is the intensity, $M$ is the magnitude on the Richter scale, and $E_0 = 10^{4.40}$ joules, which is the intensity of a very small “reference” quake. We substitute $E = 6 \times 10^{16}$ into equation (3.60) and use the properties of logarithms to obtain

$$M = \frac{2}{3} \log \frac{6 \times 10^{16}}{10^{4.40}}$$

$$= \frac{2}{3} [\log 6 \times 10^{16} - \log 10^{4.40}]$$

$$= \frac{2}{3} [\log 6 + \log 10^{16} - \log 10^{4.40}]$$

$$= \frac{2}{3} [\log 6 + 16 \log 10 - 4.40 \log 10]$$

$$= \frac{2}{3} [\log 6 + 16 - 4.40]$$

$$= \frac{2}{3} [11.6 + 0.7782]$$

$$\approx 8.25.$$

It should be noted that this earthquake is 1,000 times more intense than an earthquake of magnitude 6.25. In general, an increase of two in the magnitude corresponds to a thousand-fold increase in the intensity. □

3.6 Derivatives of Logarithmic and Exponential Functions

Now that we have learned about exponential and logarithmic functions, we can use their properties to compute their derivatives.
3.6.1 Derivatives of Logarithmic Functions

To compute the derivative of a general logarithmic function \( f(x) = \log_a x \), we use the definition of the derivative and obtain

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

\[
 = \lim_{h \to 0} \frac{1}{h} \log_a \left( \frac{x+h}{x} \right) 
\]

\[
 = \lim_{h \to 0} \frac{x}{hx} \log_a \left( \frac{x+h}{x} \right) 
\]

\[
 = \frac{1}{x} \lim_{h \to 0} \frac{x}{h} \log_a \left( \frac{x+h}{x} \right) 
\]

\[
 = \frac{1}{x} \lim_{h \to 0} \log_a \left(1 + \frac{h}{x}\right)^{x/h} 
\]

\[
 = \frac{1}{x} \log_a \left[ \lim_{h \to 0} \left(1 + \frac{h}{x}\right)^{1/t} \right], \quad t = h/x 
\]

\[
 = \frac{1}{x} \log_a e 
\]

\[
 = \frac{1}{x} \ln e 
\]

\[
 = \frac{1}{x \ln a} 
\]

We conclude with the differentiation rule

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

In the case of the natural logarithmic function, with \( a = e \), the rule simplifies to

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

From the chain rule, we obtain the more general rule

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

where \( g(x) > 0 \). Applying this rule with \( g(x) = |x| \) yields the differentiation rule

\[
 \frac{d}{dx} \ln |x| = \frac{1}{x}. 
\]
3.6. **DERIVATIVES OF LOGARITHMIC AND EXPONENTIAL FUNCTIONS**

**Example 3.35** Compute the derivative of \( f(x) = \ln(\sin x) \sin(\ln x) \).

**Solution** Using the Product Rule and Chain Rule, we obtain

\[
f'(x) = \sin(\ln x)\left[\ln(\sin x)' + \ln(\sin x)[\sin(\ln x)]'ight]
= \sin(\ln x) \frac{1}{\sin x} \cos x + \ln(\sin x) \cos(\ln x) \frac{1}{x}
= \sin(\ln x) \cot x + \frac{\ln(\sin x) \cos(\ln x)}{x}.
\]

\(\Box\)

**Example 3.36** Differentiate \( f(x) = \ln(x^e + \ln x) \).

**Solution** We have

\[
f'(x) = \frac{1}{x^e + \ln x} (x^e + \ln x)' = \frac{1}{x^e + \ln x} \left( ex^{e-1} + \frac{1}{x} \right).
\]

\(\Box\)

**Example 3.37** Differentiate \( f(x) = \ln(1 + x^2) \).

**Solution** From the Chain Rule, we have

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

\(\Box\)

**Example 3.38** Differentiate \( f(x) = \log_2(\sin x) \sin(\log_2 x) \).

**Solution** Using the Product Rule, the Chain Rule, and the rule \( (\log_a x)' = 1/(x \ln a) \), we obtain

\[
f'(x) = \sin(\log_2 x)(\log_2(\sin x)') + \log_2(\sin x)(\sin(\log_2 x))'
= \sin(\log_2 x) \frac{1}{\sin x} \cos x + \log_2(\sin x) \cos(\log_2 x) \frac{1}{x \ln 2}
= \frac{\sin(\log_2 x) \cos x}{(\sin x) \ln 2} + \frac{\log_2(\sin x) \cos(\log_2 x)}{x \ln 2}.
\]

\(\Box\)
3.6.2 Logarithmic Differentiation

The natural logarithmic function can be used to more easily differentiate functions that are complicated products or quotients, since logarithms of products and quotients are equal to sums and differences. For example, if \( f(x) = g(x)h(x) \), then

\[
\ln f(x) = \ln g(x) + \ln h(x).
\]

Differentiating both sides, we obtain

\[
\frac{f'(x)}{f(x)} = \frac{g'(x)}{g(x)} + \frac{h'(x)}{h(x)}
\]

which can then be solved for \( f'(x) \). This technique is known as logarithmic differentiation.

**Example 3.39** Given \( y = (\sin x)^{x^2+3x} \), use logarithmic differentiation to compute \( y' \).

**Solution** Taking the natural logarithm of both sides of the above equation, we have

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

Differentiating both sides implicitly with respect to \( x \) (i.e., treating \( y \) as a function of \( x \)) yields, via the Product Rule and the Chain Rule,

\[
\frac{1}{y} y' = \ln(\sin x)(x^2+3x)' + (x^2+3x)(\ln(\sin x))' = \ln(\sin x)(2x+3) + (x^2+3x) \frac{1}{\sin x} \cos x.
\]

Multiplying through by \( y \), we obtain

\[
y' = y \left[ \ln(\sin x)(2x+3) + (x^2+3x) \frac{\cos x}{\sin x} \right] = (\sin x)^{x^2+3x} \left[ \ln(\sin x)(2x+3) + (x^2+3x) \cot x \right].
\]

\( \square \)

**Example 3.40** Compute the slope of the tangent line to the graph of \( y = x^x \) at \( x = e \).

**Solution** To obtain the slope of the tangent line, we need to compute the derivative of \( x^x \) at \( x = e \). Taking the natural logarithm of both sides of the equation \( y = x^x \) we obtain

\[
\ln y = \ln x^x = x \ln x.
\]
3.6. **DERIVATIVES OF LOGARITHMIC AND EXPONENTIAL FUNCTIONS**

Differentiating both sides with respect to $x$ (and remembering that $y$ depends on $x$) yields, via the Product Rule,

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

and therefore

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

Substituting $x = 1$ into this equation yields $y' = 1^{1}(\ln 1 + 1) = 1$, and therefore the slope of the tangent line at the point $x = 1$ is equal to 1. \(\square\)

### 3.6.3 Derivatives of Exponential Functions

It is interesting to note that the derivative of $\ln x$ is $1/x$, while the derivative of an inverse function $f^{-1}(x)$ is equal to $1/f'(f^{-1}(x))$. We know the inverse function of $f(x) = \ln x$ is $f^{-1}(x) = e^x$. Then, because $f'(x) = 1/x$, it follows that

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

In other words, the derivative of $e^x$ is equal to itself.

The simplicity of this differentiation rule leads to widespread usage of $e^x$ in the solution of differential equations. For example, the solution $y(t)$ of the simple differential equation

\[
\frac{dy}{dt} = ky, \quad y(0) = y_0,
\]

where $k$ is a constant, is $y(t) = y_0e^{kt}$.

From the chain rule, we obtain the more general differentiation rule

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

This rule can be used to compute the derivative of a general exponential function $a^x$. We have

\[
\frac{d}{dx}[a^x] = \frac{d}{dx}[e^{x\ln a}] = e^{x\ln a} \ln a = a^x \ln a.
\]

We see that the derivative of any exponential function is proportional to itself, but when the base $a$ is equal to $e$, the constant of proportionality is equal to one, because $\ln e = 1$. 


In Section 2.5, we learned the power rule for differentiation,

\[ \frac{d}{dx}(x^n) = nx^{n-1}, \]

where \( n \) is any integer. Using the definition of exponentiation, we can write \( x^n = e^{n \ln x} \) and use the Chain Rule to easily prove that the power rule actually holds for any real number \( n \).

**Example 3.41** Differentiate \( f(x) = e^{x^3} \).

**Solution** Using the rule \( (e^x)' = e^x \) and the Chain Rule, we obtain

\[ f'(x) = e^{x^3}(x^3)' = e^{x^3}(3x^2). \]

\[ \square \]

**Example 3.42** Differentiate \( f(x) = e^{x^2+x} \).

**Solution** Using the Chain Rule, we obtain

\[ f'(x) = e^{x^2+x}(x^2 + x)' = e^{x^2+x}(2x + 1). \]

Another approach is to use the laws of exponents to write

\[ f(x) = e^{x^2+x} = e^{x^2}e^x \]

and use the Product Rule and the Chain Rule to obtain

\[ f'(x) = e^x(e^{x^2})' + e^{x^2}(e^x)' = e^x e^{x^2} (2x) + e^{x^2} e^x = e^{x^2+x}(2x + 1). \]

\[ \square \]

**Example 3.43** Differentiate \( f(x) = \sin(e^x) \).

**Solution** Using the Chain Rule yields

\[ f'(x) = \cos(e^x)(e^x)' = \cos(e^x)e^x. \]

\[ \square \]

**Example 3.44** Differentiate \( f(x) = \sqrt{1 + (e^{-x})^2} \).
3.6. **DERIVATIVES OF LOGARITHMIC AND EXPONENTIAL FUNCTIONS**

**Solution** Using the Chain Rule twice, we obtain

\[
\frac{d}{dx} \left[ (1 + (e^{-x})^2)^{1/2} \right] = \frac{1}{2} (1 + (e^{-x})^2)^{-1/2} (1 + (e^{-x})^2)'
\]

\[
= \frac{1}{2} (1 + (e^{-x})^2)^{-1/2} (2e^{-x})(-e^{-x})
\]

\[
= \frac{2e^{-x}(-e^{-x})}{\sqrt{1 + (e^{-x})^2}}
\]

\[
= -\frac{2e^{-2x}}{\sqrt{1 + (e^{-x})^2}}.
\]

Alternatively, we can use the laws of exponents to write \( f(x) = \sqrt{1 + e^{-2x}} \) and proceed as follows:

\[
\frac{d}{dx} \left[ (1 + (e^{-x})^2)^{1/2} \right] = \frac{1}{2} (1 + e^{-2x})^{-1/2} (1 + e^{-2x})'
\]

\[
= \frac{1}{2} (1 + e^{-2x})^{-1/2} (-2e^{-2x})
\]

\[
= (1 + e^{-2x})^{-1/2} (-e^{-2x})
\]

\[
= -\frac{2e^{-2x}}{\sqrt{1 + e^{-2x}}}.
\]

\[\square\]

**Example 3.45** Differentiate

\[
f(x) = \ln \left( \frac{e^{x^2}}{x^2 + x + 1} \right).
\]

**Solution** Using the Chain Rule and Quotient Rule, we obtain

\[
\frac{d}{dx} \left( \frac{e^{x^2}}{x^2 + x + 1} \right) = \left( \frac{e^{x^2}}{x^2 + x + 1} \right)^{-1} \left( \frac{e^{x^2}}{x^2 + x + 1} \right)'
\]

\[
= \frac{x^2 + x + 1 (x^2 + x + 1)(e^{x^2})' - e^{x^2}(x^2 + x + 1)'}{(x^2 + x + 1)^2}
\]

\[
= \frac{x^2 + x + 1 (x^2 + x + 1)e^{x^2} (2x) - e^{x^2} (2x + 1)}{(x^2 + x + 1)^2}
\]
Another way of obtaining the derivative is to use the laws of logarithms in advance. Writing

\[ f(x) = \ln e^{x^2} - \ln (x^2 + x + 1) = x^2 \ln e - \ln (x^2 + x + 1) = x^2 - \ln (x^2 + x + 1). \]

we obtain

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

\[ \square \]

**Example 3.46** Differentiate \( f(x) = \log_3 (e^x + \log_2 x) \).

**Solution** Using the Chain Rule, we obtain

\[
 f'(x) = \frac{1}{(e^x + \log_2 x) \ln 3} (e^x + \log_2 x)' \\
 = \frac{1}{(e^x + \log_2 x) \ln 3} \left( e^x + \frac{1}{x \ln 2} \right).
\]

\[ \square \]

### 3.7 Exponential Growth and Decay

In many applications, it is necessary to solve the problem of finding the function \( y(t) \) that satisfies the simple differential equation

\[
 \frac{dy}{dt} = ky,
\]

where \( k \) is a known constant, and the initial condition

\[ y(0) = y_0, \]

where \( y_0 \) is a given initial value. Together, the differential equation and initial condition specify an initial value problem. In words, the initial value problem states the following:

- The unknown quantity \( y(t) \) initially has the value \( y_0 \).
3.7. EXPONENTIAL GROWTH AND DECAY

- At any time \( t \), its rate of change with respect to time is proportional to its value at the time \( t \), with \( k \) being the constant of proportionality.

We can solve the initial value problem using the natural logarithmic and exponential functions. If we isolate the dependent variable \( y \) on one side of the differential equation and \( t \) on the other side, we obtain

\[
\frac{1}{y} \frac{dy}{dt} = k,
\]

and undoing differentiation on both sides, or anti-differentiating, yields

\[
\ln |y| = kt + C,
\]

where \( C \) is an arbitrary constant.

We then exponentiate both sides. From the cancellation equations and the laws of exponents, it follows that

\[
|y| = e^{kt} e^C,
\]

or

\[
y = Ae^{kt}
\]

where \( A \) is an arbitrary constant, that could be negative or zero. If we substitute \( t = 0 \), then we find that we must have \( A = y_0 \) and we conclude that the solution to the initial value problem is

\[
y(t) = y_0 e^{kt}.
\]

If \( k > 0 \), then \( y(t) \) grows rapidly in absolute value over time, and the differential equation is called the law of natural growth. In this case, the equation can be used as a (simplified) mathematical model for population growth, for example. If \( k < 0 \), then \( y(t) \) rapidly decays to zero over time, and the differential equation is known as the law of natural decay. This law describes, for instance, the process of radioactive decay. Finally, if \( k = 0 \), then there is no growth or decay, and \( y(t) \) remains constant for all time.

3.7.1 Population Growth

The differential equation

\[
\frac{dP}{dt} = kP
\]

can be used to model the growth of a population, denoted by \( P(t) \). The proportionality constant \( k \), also known as the relative growth rate, can be
determined experimentally by birth rates and death rates, since the number
of births and deaths per unit of time tend to be proportional to the current
population. Therefore, if the birth rate was determined to be some value $B$
and the death rate was determined to be $D$, then one model for population
growth could be
\[
\frac{dP}{dt} = (B - D)P,
\]
which would imply $P(t) = P_0 e^{(B-D)t}$, where $t = 0$ denotes the current time.
If the population of a region is being measured, then other factors, such as
immigration and emigration due to economic forces, could also be measured
and included in the determination of the relative growth rate.

3.7.2 Radioactive Decay
The mass $m(t)$ of a radioactive substance satisfies
\[
\frac{dm}{dt} = km,
\]
where $k < 0$. By measuring $m(t)$ at different times, the rate of decay,
denoted by $-k$, can be determined experimentally, since $m(t) = m_0 e^{kt}$,
where $m_0$ is the initial mass. Then, this rate can be used to determine how
much of the substance will remain after any given length of time.

3.7.3 Continuously Compounded Interest
Let $A(t)$ be the value of an investment at time $t$. With continuous com-
pounding of interest, the rate of increase of the value is proportional to its
size. It follows that
\[
\frac{dA}{dt} = rA,
\]
where $r$ is the interest rate. The value is therefore given by $A(t) = A_0 e^{rt}$,
where $A_0$ is the initial value of the investment.

Example 3.47 In 1900, the world population was approximately 1650 mil-
lion. In 1910, the population had grown to approximately 1750 million.
Determine the relative growth rate of the population.

Solution The population at time $t$, where $t$ is measured in years since 1900,
is denoted by a function $P(t)$ that satisfies the differential equation
\[
\frac{dP}{dt} = kP,
\]
where $k$ is the relative growth rate we are seeking, and the initial condition $P(0) = 1650$. This initial value problem has the solution $P(t) = 1650e^{kt}$. Substituting $t = 10$ into $P(t)$ yields

$$1650e^{10k} = 1750$$

which can be solved for $k$ using the cancellation equation $\ln e^x = x$:

$$e^{10k} = \frac{1750}{1650} \Rightarrow 10k = \ln \frac{1750}{1650} \Rightarrow k = \frac{1}{10} \ln \frac{1750}{1650} \approx 0.005884.$$

Example 3.48 The half-life of Radium-226 is 1590 years. Determine the relative decay rate.

Solution The half-life of a radioactive substance is the amount of time that is necessary for a sample of a substance to lose half of its original mass. If $m(t)$ denotes the mass of a sample of Radium-226 $t$ years after the initial time $t = 0$, then $m(t)$ satisfies the differential equation

$$\frac{dm}{dt} = km$$

where $k$ is a negative constant, and $|k|$ denotes the relative decay rate we are seeking. The function $m(t)$ also satisfies the initial condition $m(0) = m_0$, where $m_0$ is the initial mass. The solution of this initial value problem is $m(t) = m_0 e^{kt}$. Substituting $t = 1590$, it follows from the definition of half-life that

$$m_0 e^{1590k} = \frac{m_0}{2}$$

or

$$e^{1590k} = \frac{1}{2}.$$

Using the cancellation equation $\ln e^x = x$ and the laws of logarithms, we obtain

$$1590k = \ln \frac{1}{2} = \ln 1 - \ln 2 = -\ln 2.$$

We conclude that the relative decay rate is

$$|k| = \left| -\frac{\ln 2}{1590} \right| \approx 4.3594 \times 10^{-4}.$$
Example 3.49 How much of a 100 g sample of Radium-226 will remain after 1000 years?

Solution From the previous example, we have

\[ m(t) = m_0e^{kt} = 100e^{-(\ln 2)t/1590}, \]

since the initial mass \( m_0 \) is given to be 100 g. Using the definition of the general exponential function \( a^x = e^{x\ln a} \), we can rewrite \( m(t) \) as follows:

\[ m(t) = 100e^{(-t/1590)\ln 2} = 100 \cdot 2^{-t/1590}. \]

Substituting \( t = 1000 \) yields \( m(1000) \approx 64.6655 \) g.

Example 3.50 How long will it take a 100 g sample of Radium-226 to decay to 30 g?

Solution We need to find the value of \( t \) such that

\[ m(t) = 100 \cdot 2^{-t/1590} = 30. \]

Taking the logarithm to base 2 of both sides and using the cancellation equation \( \log_2 2^x = x \), we obtain

\[ -\frac{t}{1590} = \log_2 \frac{30}{100} \]

which yields

\[ t = -1590 \log_2 \frac{3}{10} \approx 2761.7752 \text{ years}. \]

3.8 Inverse Trigonometric Functions

We now wish to define inverses of the trigonometric functions, starting with the sine function \( \sin x \). Unfortunately, \( \sin x \) is not one-to-one on the domain \( -\infty < x < \infty \), as it oscillates between -1 and 1. However, we can define an inverse function if we restrict ourselves to the domain \( [-\pi/2, \pi/2] \), on which \( \sin x \) is one-to-one. We therefore define the inverse sine function, written \( \sin^{-1} x \), as follows:

\[ \sin^{-1} x = y \iff \sin y = x, \quad -\frac{\pi}{2} \leq y \leq \frac{\pi}{2}. \]
The sine and inverse sine functions satisfy the cancellation equations
\[
\sin^{-1}(\sin x) = x, \quad -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}
\]
and
\[
\sin(\sin^{-1} x) = x, \quad -1 \leq x \leq 1.
\]

We can use the cancellation equations and the Chain Rule to compute the derivative of \( y = \sin^{-1} x \). Since \( \sin y = x \), it follows that
\[
\cos y \frac{dy}{dx} = 1,
\]
which yields
\[
\frac{dy}{dx} = \frac{1}{\cos y}.
\]
Since \( \cos y \geq 0 \) for \( -\pi/2 \leq y \leq \pi/2 \), we can use the identity \( \sin^2 y + \cos^2 y = 1 \) to conclude
\[
\cos y = \sqrt{1 - \sin^2 y},
\]
and then use the relation \( x = \sin y \) to obtain
\[
\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1 - x^2}}, \quad -1 < x < 1.
\]

The inverse cosine function, written \( \cos^{-1} x \), is defined similarly. The cosine function is one-to-one on the domain \([0, \pi]\), so the inverse cosine is defined by
\[
\cos^{-1} x = y \iff \cos y = x, \quad 0 \leq y \leq \pi.
\]
Using the same approach as with \( \sin^{-1} x \), we can easily compute its derivative:
\[
\frac{d}{dx}(\cos^{-1} x) = -\frac{1}{\sqrt{1 - x^2}}, \quad -1 < x < 1.
\]

The tangent function \( \tan x \) is one-to-one on the domain \((-\pi/2, \pi/2)\). Note that we must use the open interval in this case, since \( \tan x \) has vertical asymptotes at \( x = \pm \pi/2 \). The inverse tangent function \( \tan^{-1} x \) is therefore defined as follows:
\[
\tan^{-1} x = y \iff \tan y = x, \quad -\frac{\pi}{2} < y < \frac{\pi}{2}.
\]
The vertical asymptotes of \( \tan x \) translate to horizontal asymptotes in \( \tan^{-1} x \); specifically, we have
\[
\lim_{x \to \infty} \tan^{-1} x = \frac{\pi}{2}, \quad \lim_{x \to -\infty} \tan^{-1} x = -\frac{\pi}{2}.
\]
To obtain the derivative of $\tan^{-1} x$, we proceed as with $\sin^{-1} x$. Differentiating both sides of the relation $\tan y = x$ with respect to $x$ yields

$$\sec^2 y \frac{dy}{dx} = 1,$$

and it follows from the identity $\sec^2 y = 1 + \tan^2 y$ that

$$\frac{dy}{dx} = \frac{1}{1 + \tan^2 y}.$$

Since $\tan y = x$, we can conclude that

$$\frac{d}{dx} (\tan^{-1} x) = \frac{1}{1 + x^2}, \quad -1 < x < 1.$$

**Example 3.51** Simplify the expression $\sin(\tan^{-1} x)$.

**Solution** There are two approaches we can use. The first approach, an algebraic one, uses the fact that $\tan x = \sin x / \cos x$ and $\sec x = 1 / \cos x$, as well as the identity $\sec^2 x = 1 + \tan^2 x$. We have

$$\sin(\tan^{-1} x) = \frac{\sin(\tan^{-1} x)}{\cos(\tan^{-1} x)} \cos(\tan^{-1} x)$$

$$= \tan(\tan^{-1} x) \frac{1}{\sec(\tan^{-1} x)}$$

$$= \frac{x}{\sqrt{\sec^2(\tan^{-1} x)}}$$

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

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

The second approach is geometric. Consider a right triangle where one of the angles is equal to $\theta = \tan^{-1} x$. Then, from the cancellation equation $\tan(\tan^{-1} x) = x$, we have $\tan \theta = x$. Since $\tan \theta$ is the ratio of the length of the side opposite the angle $\theta$ to the length of the adjacent side, we can assume that the length of the opposite side is $x$ and the length of the adjacent side is 1. Since $\sin(\tan^{-1} x) = \sin \theta$, and $\sin \theta$ is defined to be the ratio of the length of the opposite side to the length of the hypotenuse, we simply need to compute this ratio. The opposite side has length $x$, and since the adjacent side has length 1, it follows that the hypotenuse has length $\sqrt{1 + x^2}$. Therefore

$$\sin(\tan^{-1} x) = \sin \theta = \frac{x}{\sqrt{1 + x^2}}.$$

$\square$
Example 3.52 Simplify the expression $\tan(\sin^{-1} x)$.

**Solution** Proceeding as in the last example, we let $\theta = \sin^{-1} x$, so by the cancellation equation $\sin(\sin^{-1} x) = x$, we have $\sin \theta = x$. Since sine is the ratio of opposite to hypotenuse, we can choose the opposite side to have length $x$ and the hypotenuse to have length 1. Since tangent is equal to ratio of opposite to adjacent, and the adjacent side must have length $\sqrt{1 - x^2}$, we have

$$\tan(\sin^{-1} x) = \tan \theta = \frac{x}{\sqrt{1 - x^2}}.$$  

Another approach is to use the fact that $\tan \theta = \sin \theta / \cos \theta$ and use the cancellation $\sin(\sin^{-1} x) = x$ to obtain

$$\tan(\sin^{-1} x) = \frac{\sin(\sin^{-1} x)}{\cos(\sin^{-1} x)} = \frac{x}{\cos(\sin^{-1} x)}.$$  

We can then use the identity $\cos^2 \theta = 1 - \sin^2 \theta$ to obtain

$$\tan(\sin^{-1} x) = \frac{x}{\cos(\sin^{-1} x)} = \frac{x}{\sqrt{1 - [\sin(\sin^{-1} x)]^2}} = \frac{x}{\sqrt{1 - x^2}}.$$  

□

Example 3.53 Consider the diagram in Figure 3.6. Where should the point $P$ be placed so as to maximize the angle $\theta$?

**Solution** Let $x$ be the distance along the base from the left edge to the point $P$. Then, because the tangent is defined to be the ratio of the length of the opposite side to the length of the adjacent side, it follows that

$$\tan \alpha = \frac{5}{x}.$$  

Similarly, we have

$$\tan \beta = \frac{2}{3 - x}.$$  

Because $\theta$ is maximized when the angles $\alpha$ and $\beta$ are minimized, we can determine the correct placement of $P$ by finding the value of $x$ such that

$$y = \alpha + \beta = \tan^{-1} \frac{5}{x} + \tan^{-1} \frac{2}{3 - x}$$

is a minimum.
To find the minimum, we must determine where $y' = 0$. Differentiating with respect to $x$, we obtain

$$y' = \frac{1}{1 + \left(\frac{5}{x}\right)^2} \left(\frac{5}{x}\right)' + \frac{1}{1 + \left(\frac{2}{3 - x}\right)^2} \left(\frac{2}{3 - x}\right)'$$

$$= \frac{1}{1 + \frac{25}{x^2}} \left(-\frac{5}{x^2}\right) + \frac{1}{1 + \frac{4}{(3 - x)^2}} \frac{2}{(3 - x)^2}$$

$$= -\frac{5}{x^2 + 25} + \frac{2}{(3 - x)^2 + 4}.$$

Setting $y' = 0$ yields the equation

$$\frac{2}{(3 - x)^2 + 4} = \frac{5}{x^2 + 25}. $$
Cross-multiplying, we obtain the simper equation
\[ 2(x^2 + 25) = 5[(3 - x)^2 + 4] \]
which, upon expanding, yields
\[ 2x^2 + 50 = 5(9 - 6x + x^2) + 20 = 5x^2 - 30x + 65 \]
which then simplifies to
\[ 3x^2 - 30x + 15 = 0 \]
or
\[ x^2 - 10x + 5 = 0. \]
Applying the quadratic formula, we obtain
\[ x = \frac{10 \pm \sqrt{100 - 4(5)(1)}}{2} = 5 \pm \sqrt{80} = 5 \pm 2\sqrt{5}. \]
Since we must have 0 < x < 3, we must choose the negative square root and therefore the critical point \( x_c \) of \( y \) occurs at
\[ x_c = 5 - 2\sqrt{5} \approx 0.5279. \]
To be certain that this critical point corresponds to a minimum, we must compute \( y'' \). We have
\[ y'' = \frac{5}{(x^2 + 25)^2}(2x) - \frac{2}{((3 - x)^2 + 4)^2} (2(3-x)(-1)) = \frac{10x}{(x^2 + 25)^2} + \frac{4(3-x)}{((3 - x)^2 + 4)^2}. \]
Since the critical point \( x_c \) lies between 0 and 3, it is clear that substituting \( x_c \) for \( x \) in \( y'' \) yields a positive value and therefore it corresponds to a minimum.

\[ \square \]

### 3.9 Indeterminate Forms and l’Hospital’s Rule

Suppose that \( f(x) \) and \( g(x) \) are two differentiable functions such that \( f(a) = g(a) = 0 \), and both derivatives are continuous at \( x = a \). Then, it is not clear whether
\[ L = \lim_{x \to a} \frac{f(x)}{g(x)} \]
exists, since it depends on the rates at which \( f(x) \) and \( g(x) \) approach zero as \( x \) approaches \( a \). However, because both functions are differentiable, it follows that

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

In summary, we can determine the limit \( L \) of \( f/g \) as \( x \to a \) by differentiating both functions and computing the limit of \( f'/g' \). If \( f'(a) = g'(a) = 0 \) and both derivatives are themselves continuously differentiable at \( a \), we can repeat this process until the limit can be determined.

This technique of computing the limit of the quotient of the derivatives is known as l’Hospital’s Rule. The rule states that

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

under the following conditions:

- \( f \) and \( g \) are differentiable at \( a \), and \( g'(x) \neq 0 \) for \( x \) near \( a \), except possibly at \( x = a \) itself.

- Either

\[
\lim_{x \to a} f(x) = 0 \quad \text{and} \quad \lim_{x \to a} g(x) = 0,
\]

that is, \( L \) is an indeterminate form of type \( 0/0 \), or

\[
\lim_{x \to a} f(x) = \pm \infty \quad \text{and} \quad \lim_{x \to a} g(x) = \pm \infty,
\]

in which case \( L \) is an indeterminate form of type \( \infty/\infty \).
It should be noted that this rule applies whether the limit $L$ itself is either finite or infinite.

Other indeterminate forms can be evaluated using l’Hopital’s Rule. For example, suppose that as $x \to a$, $f(x) \to 0$ and $g(x) \to \infty$. Then, the limit

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

is an indeterminate form of type $0 \cdot \infty$. By rewriting the product as a quotient $f/(1/g)$ or $g/(1/f)$, we obtain an indeterminate form of type $0/0$ or $\infty/\infty$, in which case l’Hopital’s Rule can be applied.

Similarly, if $f(x)$ and $g(x)$ both become infinite as $x \to a$, then the limit

$$L = \lim_{x \to a} [f(x) - g(x)]$$

is an indeterminate form of type $\infty - \infty$. This difference can be converted into a quotient using techniques such as computing a common denominator, rationalizing or removing a common factor. The result is a new indeterminate form for which l’Hospital’s Rule is useful.

Finally, the limit

$$L = \lim_{x \to a} [f(x)]^{g(x)}$$

can be an indeterminate form of type $0^0$, $\infty^0$, or $1^\infty$, depending on the limits of $f(x)$ and $g(x)$ as $x \to a$. In all three cases, we can compute $L$ by computing the limit of $\ln[f(x)]^{g(x)} = g(x) \ln f(x)$, which is an indeterminate form of type $0 \cdot \infty$.

Example 3.54 Compute

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

Solution This limit is an indeterminate form of type $0/0$. Applying l’Hospital’s Rule yields

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

Example 3.55 Compute

$$\lim_{x \to \infty} \frac{x}{x^2 + 1}.$$

Solution This limit is an indeterminate form of type $\infty/\infty$. Applying l’Hospital’s Rule yields

$$\lim_{x \to \infty} \frac{x}{x^2 + 1} = \lim_{x \to \infty} \frac{1}{2x} = 0.$$
Example 3.56 Compute

\[ \lim_{x \to \infty} \frac{x}{e^x}. \]

**Solution** This limit is an indeterminate form of type \( \infty/\infty \). Applying 
l'Hospital's Rule yields

\[ \lim_{x \to \infty} \frac{x}{e^x} = \lim_{x \to \infty} \frac{1}{e^x} = 0. \]

\[ \blacksquare \]

Example 3.57 Compute

\[ \lim_{x \to \infty} \frac{x}{(x^2 + 1)^{1/2}}. \]

**Solution** This limit is an indeterminate form of type \( \infty/\infty \). Applying 
l'Hospital's Rule yields

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

This limit is also an indeterminate form of type \( \infty/\infty \). Applying l'Hospital’s 
Rule again yields

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

Unfortunately, we are back where we started, so we cannot evaluate the limit 
by applying l'Hospital’s Rule directly. We can instead use the fact that

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

to conclude that

\[ \lim_{x \to \infty} \frac{x}{(x^2 + 1)^{1/2}} = \left( \lim_{x \to \infty} \frac{x^2}{x^2 + 1} \right)^{1/2}. \]

The limit on the right side is an indeterminate form of type \( \infty/\infty \). Applying 
l'Hospital’s Rule yields

\[ \lim_{x \to \infty} \frac{x^2}{x^2 + 1} = \lim_{x \to \infty} \frac{2x}{2x} = 1. \]
It follows that
\[
\lim_{x \to \infty} \frac{x}{(x^2 + 1)^{1/2}} = \left( \lim_{x \to \infty} \frac{x^2}{(x^2 + 1)^{1/2}} \right)^{1/2} = 1^{1/2} = 1.
\]

\[
\square
\]

**Example 3.58** Compute
\[
\lim_{x \to 0} (\sin x)(\ln x).
\]

**Solution** This limit is an indeterminate form of type \(0 \cdot \infty\). We can write it as a quotient using \(\sin x = 1 / \csc x\) and obtain
\[
\lim_{x \to 0} \frac{\ln x}{\csc x}.
\]
This limit is an indeterminate form of type \(\infty / \infty\). Applying l'Hospital’s Rule yields
\[
\lim_{x \to 0} \frac{\ln x}{\csc x} = \lim_{x \to 0} \frac{1}{-\csc x \cot x} = \lim_{x \to 0} -\sin^2 x / x \cos x,
\]
using the fact that \(\cot x = \cos x / \sin x\). This limit is an indeterminate form of type \(0/0\). Applying l'Hospital’s Rule again, in conjunction with the Product Rule, yields
\[
\lim_{x \to 0} -\frac{\sin^2 x}{x \cos x} = \lim_{x \to 0} \frac{-2 \sin x \cos x}{\cos x - x \sin x} = 0.
\]
\[
\square
\]

**Example 3.59** Compute
\[
\lim_{x \to 1} \frac{1}{\ln x} - \frac{1}{x - 1}.
\]

**Solution** This limit is an indeterminate form of type \(\infty - \infty\). Finding a common denominator, we can rewrite this limit as
\[
\lim_{x \to 1} \frac{x - 1 - \ln x}{(\ln x)(x - 1)}.
\]
This limit is an indeterminate form of type \(0/0\). Applying l'Hospital’s Rule yields
\[
\lim_{x \to 1} \frac{x - 1 - \ln x}{(\ln x)(x - 1)} = \lim_{x \to 1} \frac{1 - 1/x}{(x - 1)/x + \ln x}.
\]
CHAPTER 3. INVERSE FUNCTIONS

Multiplying the numerator and denominator by \( x \) yields

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

This limit is an indeterminate form of type 0/0. Applying l’Hospital’s Rule again yields

\[
\lim_{x \to 1} \frac{x - 1}{x - 1 + \ln x} = \lim_{x \to 1} \frac{1}{1 + \ln x + x (1/x) \ln x} = \frac{1}{2}.
\]

\( \square \)

**Example 3.60** Compute

\[
\lim_{x \to 0} (1 + x)^{1/x}.
\]

**Solution** This limit is an indeterminate form of type 1\( ^\infty \). We write

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

where

\[
L = \lim_{x \to 0} \ln(1 + x)^{1/x} = \lim_{x \to 0} \frac{1}{x} \ln(1 + x).
\]

This limit is an indeterminate form of type 0/0. Applying l’Hospital’s Rule yields

\[
L = \lim_{x \to 0} \frac{\ln(1 + x)}{x} = \lim_{x \to 0} \frac{1/(1 + x)}{1} = 1.
\]

It follows that

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

\( \square \)

**Example 3.61** Compute

\[
\lim_{x \to \infty} (1 + x)^{1/x}.
\]

This limit is an indeterminate form of type \( \infty^0 \). We write

\[
\lim_{x \to \infty} (1 + x)^{1/x} = e^L
\]

where

\[
L = \lim_{x \to \infty} \ln(1 + x)^{1/x} = \lim_{x \to \infty} \frac{1}{x} \ln(1 + x).
\]
This limit is an indeterminate form of type $\infty/\infty$. Applying l’Hospital’s Rule yields

\[
L = \lim_{x \to \infty} \frac{\ln(1 + x)}{x} = \lim_{x \to \infty} \frac{1/(1 + x)}{1} = \lim_{x \to \infty} \frac{1}{1 + x} = 0.
\]

It follows that

\[
\lim_{x \to \infty} (1 + x)^{1/x} = e^L = e^0 = 1.
\]
Chapter 4

Applications of Derivatives

4.1 Maximum and Minimum Values

In many applications, it is necessary to determine where a given function, called an objective function, attains its minimum or maximum value. For example, a business wishes to maximize profit, so it can construct a function that relates its profit to variables such as payroll or maintenance costs.

The problem of determining the maximum or minimum value of an objective function is called optimization, since the minimum or maximum value typically corresponds to some sort of optimal condition. We now consider the basic problem of finding a maximum or minimum value of a general function \( f(x) \) that depends on a single independent variable \( x \). First, we must precisely define what it means for a function to have a maximum or minimum value.

**Definition 4.1 (Absolute extrema)** A function \( f \) has a **absolute maximum** or **global maximum** at \( c \) if \( f(c) \geq f(x) \) for all \( x \) in the domain of \( f \). The number \( f(c) \) is called the **maximum value** of \( f \) on its domain. Similarly, \( f \) has a **absolute minimum** or **global minimum** at \( c \) if \( f(c) \leq f(x) \) for all \( x \) in the domain of \( f \). The number \( f(c) \) is then called the **minimum value** of \( f \) on its domain. The maximum and minimum values of \( f \) are called the **extreme values** of \( f \), and the absolute maximum and minimum are each called an **extremum** of \( f \).

Before computing the maximum or minimum value of a function, it is natural to ask whether it is possible to determine in advance whether a function even has a maximum or minimum, so that effort is not wasted in trying to
solve a problem that has no solution. The following result is very helpful in answering this question.

**Theorem 4.1 (Extreme Value Theorem)** If $f$ is continuous on $[a, b]$, then $f$ has an absolute maximum and an absolute minimum on $[a, b]$.

Now that we can easily determine whether a function has a maximum or minimum on a closed interval $[a, b]$, we can develop a method for actually finding them. It turns out that it is easier to find points at which $f$ attains a maximum or minimum value in a “local” sense, rather than a “global” sense. In other words, we can best find the absolute maximum or minimum of $f$ by finding points at which $f$ achieves a maximum or minimum with respect to “nearby” points, and then determine which of these points is the absolute maximum or minimum. The following definition makes this notion precise.

**Definition 4.2 (Local extrema)** A function $f$ has a **local maximum** at $c$ if $f(c) \geq f(x)$ for all $x$ in an open interval containing $c$. Similarly, $f$ has a **local minimum** at $c$ if $f(c) \leq f(x)$ for all $x$ in an open interval containing $c$. A local maximum or local minimum is also called a **local extremum**.

**Example 4.1** Figure 4.1 illustrates local maxima and minima, or, collectively, local extrema, of a function $f$ on the interval $[0, 6]$. Note that at each local maximum or minimum, the derivative of $f$ is equal to zero, or $f$ is not differentiable at all, as is the case at $x = 5$. □

The preceding example illustrates how we can characterize local maxima of a function. At each point at which $f$ has a local maximum, the function either has a horizontal tangent line, or no tangent line due to not being differentiable. It turns out that this is true in general, and a similar statement applies to local minima. To state the formal result, we first introduce the following definition, which will also be useful when describing a method for finding local extrema.

**Definition 4.3 (Critical Number)** A number $c$ in the domain of a function $f$ is a **critical number** of $f$ if $f'(c) = 0$ or $f'(c)$ does not exist.

The following result describes the relationship between critical numbers and local extrema.

**Theorem 4.2 (Fermat’s Theorem)** If $f$ has a local minimum or local maximum at $c$, then $c$ is a critical number of $f$; that is, either $f'(c) = 0$ or $f'(c)$ does not exist.
4.1. MAXIMUM AND MINIMUM VALUES

Figure 4.1: Examples of local extrema of a function \( f \) on the interval \([0, 6]\).
The function has local maxima at \( x = 1, x = 3, \) and \( x = 5 \), while there are local minima at \( x \approx 0.18, x = 2 \) and \( x \approx 3.79 \). The absolute maximum occurs at \( x = 3 \), with maximum value 5, while the absolute minimum occurs at \( x \approx 3.79 \), with approximate minimum value of 0.2986.

It is important to interpret this result carefully. It states that wherever \( f \) has a local extremum, \( f \) has a critical number. The converse of this statement, that \( f \) has a local extremum wherever \( f \) has a critical number, is not true, as the following examples illustrate.

**Example 4.2** Figure 4.2 shows the function \( f(x) = x^3 \). Because \( f'(x) = 3x^2 \), \( f \) has a critical number at \( x = 0 \). However, \( x = 0 \) is neither a local minimum nor local maximum of \( f \). □

**Example 4.3** In Figure 4.1, the function \( f \) is not differentiable at \( x = 4 \), but that point is neither a local minimum nor local maximum of \( f \). □
CHAPTER 4. APPLICATIONS OF DERIVATIVES

200

CHAPTER 4. APPLICATIONS OF DERIVATIVES

Figure 4.2: Graph of \( y = x^3 \). Note that the function has a horizontal tangent at the point \((0, 0)\), but \( x = 0 \) is neither a local minimum or local maximum.

When searching for a maximum or minimum of a function \( f \) on a closed interval \([a, b]\), it is not enough to examine all critical numbers, as the following example illustrates.

**Example 4.4** Figure 4.3 shows the graph of the function \( f(x) = (x - 1)^2 \) on the interval \([0, 3]\). The only critical number is at \( x = 1 \), which corresponds to a local minimum. There is no local maximum, but by the Extreme Value Theorem, \( f \) must have an absolute maximum on the interval. It actually occurs at the right endpoint, \( x = 3 \). Therefore, in addition to any critical numbers, we must check the endpoints in searching for the absolute maximum or minimum. \( \square \)

It can be shown that it is not only necessary, but sufficient, to examine all critical numbers of \( f \) on \([a, b]\), as well as the endpoints \( x = a \) and \( x = b \), to find the absolute maximum and minimum of \( f \) on \([a, b]\).
4.1. MAXIMUM AND MINIMUM VALUES

Figure 4.3: Graph of $y = (x - 1)^2$ on the interval $[0, 3]$. The function has no local maximum on the interval, and assumes its absolute maximum at an endpoint, $x = 3$.

Now, we are ready to describe a fairly systematic approach for finding the absolute maximum or minimum value of a continuous function $f$ on a closed interval $[a, b]$. We will refer to this method as the closed interval method. The method proceeds as follows:

1. Determine all of the critical numbers of $f$ on $[a, b]$; that is, determine every number $c$ in $[a, b]$ at which $f'(c) = 0$ or $f'(c)$ does not exist.

2. Compute the value of $f$ at each critical number.

3. Compute the value of $f$ at the endpoints of $[a, b]$.

4. The largest of the values in steps 2 and 3 is the absolute maximum value of $f$ on $[a, b]$, and the smallest of these values is the absolute minimum value.
This method serves as the cornerstone of methods for solving the optimization problem discussed at the beginning of this section. Later in this chapter we will discuss the optimization problem for the case where the objective function depends on a single variable.

**Example 4.5** Does the function \( f(x) = \frac{1}{x} \) have an absolute maximum or absolute minimum on the interval \([0, 1]\)? Explain why or why not.

**Solution** This function does not have an absolute maximum on \([0, 1]\), but it does have an absolute minimum, at \( x = 1 \). The function is not guaranteed to have both an absolute maximum and absolute minimum on \([0, 1]\) because it does not satisfy the assumptions of the Extreme Value Theorem (see Section 4.1). The theorem states that a function \( f(x) \) has an absolute maximum and absolute minimum on \([a, b]\) if it is continuous on \([a, b]\). However, \( f(x) = \frac{1}{x} \) is not continuous on \([0, 1]\), because it is undefined at \( x = 0 \); in fact, it has an vertical asymptote there. It is continuous on \((0, 1]\), but that is not sufficient to ensure the existence of an absolute maximum, which must be finite. It can be shown that a function \( f(x) \) that is continuous on a closed interval \([a, b]\) must be bounded above and below; that is, there exist values \( m \) and \( M \) such that \( m \leq f(x) \leq M \) for all \( x \) in \([a, b]\). Therefore, it must have an absolute maximum and absolute minimum on \([a, b]\).

**Example 4.6** Find the absolute maximum and absolute minimum values of \( f(x) = x^2 - 2x + 2 \) on the interval \([-1, 4]\).

**Solution** This function is continuous on \([-1, 4]\), so it is guaranteed to have an absolute maximum and absolute minimum on this interval. The only \( x \)-values at which an absolute maximum or absolute minimum can occur are at the endpoints \( x = -1 \) and \( x = 4 \), and any critical numbers of \( f(x) \) on \([-1, 4]\). To find the critical numbers, we first note that \( f \) is differentiable on \([-1, 4]\), so any critical numbers are located at points at which \( f'(x) = 0 \). We compute \( f'(x) = 2x - 2 \), and conclude that the only critical number that \( f \) has on \([-1, 4]\) is at \( x = 1 \), since \( f'(1) = 2(1) - 2 = 0 \).

We now have three “candidates” for an absolute maximum and absolute minimum: \( x = 1 \), \( x = -1 \), and \( x = 4 \). It is not possible for \( f(x) \) to assume its absolute maximum or absolute minimum value on \([-1, 4]\) at any other point. Evaluating \( f(x) \) at these three points, we obtain \( f(1) = 1 \), \( f(-1) = 5 \) and \( f(4) = 10 \). We conclude that the absolute minimum occurs at \( x = 1 \), and the absolute minimum value is 1. Also, the absolute maximum occurs at \( x = 4 \), and the absolute maximum value is 10. The function is shown in Figure 4.4. \( \Box \)
4.2. THE MEAN VALUE THEOREM

Figure 4.4: The function $f(x) = x^2 - 2x + 2$ on the interval $[-1, 4]$. The absolute maximum occurs at $x = 4$, and the absolute minimum occurs at $x = 1$.

Remark The terms “absolute maximum” and “absolute minimum” refer to $x$-values, while “absolute maximum value” and “absolute minimum value”, or simply “maximum value” and ”minimum value”, refer to the corresponding $y$-values. □

4.2 The Mean Value Theorem

In upcoming sections, we will be learning how the derivatives of a function can be used to understand the behavior of the function on its domain. While the derivative describes the behavior of a function at a point, we often need to determine a function’s behavior on an interval, as we have seen from our consideration of the problem of finding the maximum or minimum value on an interval.
Our study of how the local behavior described by the derivative influences the global behavior of the function on an interval relies largely on a single theoretical result called the Mean Value Theorem. In this section we will present the statement and proof of this theorem, as well as some of its most basic interpretations and applications. In later sections we will discover even more applications of the theorem.

Before we can state and prove the Mean Value Theorem, we need the following simpler result.

**Theorem 4.3 (Rolle’s Theorem)** If $f$ is continuous on a closed interval $[a, b]$ and is differentiable on the open interval $(a, b)$, and if $f(a) = f(b)$, then $f'(c) = 0$ for some number $c$ in $(a, b)$.

**Proof** There are two cases to consider:

- If $f(x)$ is equal to a constant on $[a, b]$, then $f'(x) = 0$ for all $x$ in $(a, b)$.

- If $f(x)$ is not constant on $[a, b]$, then we first apply the Extreme Value Theorem to conclude that $f$ has an absolute maximum and an absolute minimum on $[a, b]$. Since $f$ is not constant on $[a, b]$, there must exist some number $x$ in $(a, b)$ such that $f(x) > f(a)$ or $f(x) < f(a)$. This means that $f(a)$ cannot be both the maximum value and minimum value on $[a, b]$, and since $f(a) = f(b)$, it follows that either the absolute maximum or the absolute minimum must occur in the open interval $(a, b)$. Since an absolute maximum or an absolute minimum occurring in an open interval must also be a local maximum or local minimum, we can apply Fermat’s Theorem to conclude that there exists at least one critical number $c$ in $(a, b)$. By the definition of critical number, either $f'(c) = 0$ or $f'(c)$ does not exist. However, $f$ is assumed to be differentiable on $(a, b)$, so $f'(c)$ must exist. We conclude that $f'(c) = 0$.

\[ \square \]

**Example 4.7** In Figure 4.5, the function $f(x) = -x^2 + 2x + 1$ satisfies $f(0) = f(2)$, and has a horizontal tangent at $x = 1$. In other words, $f'(1) = 0$, as suggested by Rolle’s Theorem. \[ \square \]

We can now state and prove the Mean Value Theorem itself, as the proof will require Rolle’s Theorem.
4.2. THE MEAN VALUE THEOREM

Figure 4.5: The function \( f(x) = -x^2 + 2x + 1 \) satisfies \( f(0) = f(2) = 1 \), so by Rolle’s Theorem, it must satisfy \( f'(c) = 0 \) for some \( c \) in \( (0, 2) \). In fact, \( f'(1) = 0 \).

**Theorem 4.4** (Mean Value Theorem) If \( f \) is continuous on a closed interval \([a, b]\) and is differentiable on the open interval \((a, b)\), then

\[
\frac{f(b) - f(a)}{b - a} = f'(c) \tag{4.1}
\]

for some number \( c \) in \((a, b)\).

**Proof** Define the function \( g(x) \) by

\[
g(x) = f(x) - \frac{f(b) - f(a)}{b - a} (x - a). \tag{4.2}
\]

Then, because \( f \) is continuous on \([a, b]\) and differentiable on \((a, b)\), so is \( g \). Furthermore,

\[
g(a) = f(a) - \frac{f(b) - f(a)}{b - a} (a - a) = f(a), \tag{4.3}
\]
and
\[ g(b) = f(b) - \frac{f(b) - f(a)}{b - a}(b - a) = f(b) - [f(b) - f(a)] = f(a). \] (4.4)

Since \( g(a) = g(b) \), Rolle’s Theorem applies to \( g \), and we can conclude that there is some number \( c \) in \((a, b)\) such that \( g'(c) = 0 \). However, since
\[ g'(x) = f'(x) - \frac{d}{dx} \left[ \frac{f(b) - f(a)}{b - a} (x - a) \right] = f'(x) - \frac{f(b) - f(a)}{b - a}, \] (4.5)
we must have
\[ f'(c) - \frac{f(b) - f(a)}{b - a} = 0, \] (4.6)
or
\[ f'(c) = \frac{f(b) - f(a)}{b - a}. \] (4.7)

**Remark** The expression
\[ \frac{f(b) - f(a)}{b - a} \] (4.8)
is the slope of the secant line passing through the points \((a, f(a))\) and \((b, f(b))\). The Mean Value Theorem therefore states that under the given assumptions, the slope of this secant line is equal to the slope of the tangent line of \( f \) at the point \((c, f(c))\), where \( c \) is in \((a, b)\). This is illustrated in Figure 4.6.

The Mean Value Theorem has the following practical interpretation: the average rate of change of \( y = f(x) \) with respect to \( x \) on an interval \([a, b]\) is equal to the instantaneous rate of change \( y \) with respect to \( x \) at some point in \((a, b)\). The following example illustrates how this interpretation can be applied.

**Example 4.8** Suppose that you are driving along the freeway and you see a Highway Patrol car on the side of the road looking for speeders. You slow down in time, and the officer’s radar gun determines that you are traveling at a reasonable speed. After passing, you speed up again, and after traveling three miles within the next two minutes, you encounter another officer and slow down, only to speed up after passing once again. The second officer does not pursue you and you think you are safe, but after a few minutes you are pursued by the officer, stopped, and given a ticket for driving in excess of 90 mph.
4.2. THE MEAN VALUE THEOREM

Figure 4.6: Illustration of the Mean Value Theorem. For the function \( f(x) = x^2 - 4x + 6 \), the slope of the secant line passing through the points \((1, 3)\) and \((4, 6)\) has slope 1, as does the tangent line passing through the point \((5/2, 9/4)\). In other words, \( (f(4) - f(1))/(4-1) = f'(5/2) \).

How could it be determined that you were speeding if neither officer’s radar gun indicated that you were? The officers could collaborate and determine that you traveled three miles in two minutes, which is equivalent to 90 mph. Therefore, your average speed during those two minutes was 90 mph. By the Mean Value Theorem, your instantaneous speed must have been 90 mph at some point during that time. To see this mathematically, let \( f(t) \) be your position along the road at time \( t \), where \( t = 0 \) corresponds to the time that you passed the first officer, \( t \) is measured in hours, and \( f(t) \) indicates position in miles. Then, since two minutes is equal to 1/30th of an hour, we have

\[
\frac{f(1/30) - f(0)}{1/30 - 0} = \frac{3}{1/30} = 90, \tag{4.9}
\]
and by the Mean Value Theorem, we must have \( f'(c) = 90 \) for some \( c \) between 0 and 1/30. In this situation, your wise choice is to fight the ticket, since calculus may not necessarily be admissible in court. □

We know that if a function \( f(x) \) is equal to a constant, then \( f'(x) = 0 \). One of the most basic consequences of the Mean Value Theorem is the converse of this statement: if a function’s derivative is equal to zero on some domain, then the function must be constant on that domain. We now state this precisely and show how the Mean Value Theorem can be used to prove it.

**Theorem 4.5** If \( f'(x) = 0 \) on an open interval \((a, b)\), then \( f \) is constant on \((a, b)\).

**Proof** Let \( c \) and \( d \) be any two numbers in \((a, b)\), with \( c \neq d \). Then \( f \) is differentiable on \((c, d)\) and continuous on \([c, d]\). By the Mean Value Theorem, there is a number \( u \) in \((c, d)\) such that

\[
\frac{f(c) - f(d)}{c - d} = f'(u). \tag{4.10}
\]

However, since \( u \) is in \((c, d)\), \( u \) is in \((a, b)\), and so \( f'(u) = 0 \). Since \( c \neq d \), it follows that \( f(c) = f(d) \). Because this is true for any numbers \( c \) and \( d \) in \((a, b)\), \( f \) must be constant on \((a, b)\). □

Intuitively, this result makes sense because the statement \( f'(x) = 0 \) implies that \( y = f(x) \) is not changing with respect to \( x \) at all, and therefore \( y \) must have a constant value.

The preceding theorem yields the following basic result whose importance will become clear in the study of integral calculus.

**Theorem 4.6** If \( f'(x) = g'(x) \) on an open interval \((a, b)\), then \( f(x) = g(x) + c \), where \( c \) is a constant.

**Proof** Let \( F(x) = f(x) - g(x) \). Then \( F'(x) = 0 \) on \((a, b)\). It follows from the previous theorem that \( F \) is constant on \((a, b)\), which implies that \( f(x) - g(x) = c \) on \((a, b)\), for some constant \( c \). □

**Example 4.9** Let \( f(x) = x^2 + 2 \) and \( g(x) = x^2 - 1 \). These functions have the same derivative, as \( f'(x) = 2x \) and \( g'(x) = 2x \). However, the functions themselves differ by a constant, 3. □
Example 4.10 Show that the function

\[ f(x) = x^2 - 4x \]  

has a horizontal tangent in the interval \([0, 4]\).

Solution We first evaluate \( f(x) \) at the endpoints of the given interval, \( x = 0 \) and \( x = 4 \). This yields

\[ f(0) = 0^2 - 4(0) = 0, \quad f(4) = 4^2 - 4(4) = 16 - 16 = 0. \]

By Rolle’s Theorem, since \( f(0) = f(4) \), there must exist a point \( c \) in \([0, 4]\) such that \( f'(c) = 0 \). □

Example 4.11 Let

\[ f(x) = \sec^2 \pi x. \]  

Show that there exists a point \( c \) in the interval \([0, 1/4]\) such that \( f'(c) = 4 \).

Solution We compute the difference quotient

\[ \frac{f(b) - f(a)}{b - a}, \]  

with \( a = 0 \) and \( b = 1/4 \). Since

\[ f(0) = \sec^2 0 = \frac{1}{\cos^2 0} = 1, \]  

and

\[ f(1/4) = \sec^2 \frac{\pi}{4} = \frac{1}{\cos^2 \frac{\pi}{4}} = \frac{1}{(\sqrt{2}/2)^2} = \frac{1}{1/2} = 2, \]

we have

\[ \frac{f(1/4) - f(0)}{1/4 - 0} = \frac{2 - 1}{1/4 - 0} = \frac{1}{1/4} = 4. \]  

By the Mean Value Theorem, there must exist a point \( c \) in \([0, 1/4]\) such that \( f'(c) = 4 \). □

Example 4.12 Show that the function

\[ f(x) = x^3 + x \]  

has only one real root; that is, show that there is only one real number \( x \) such that \( f(x) = 0 \).
CHAPTER 4. APPLICATIONS OF DERIVATIVES

Solution We prove this by contradiction. Suppose that \( f(x) \) has more than one real root. This implies that there are at least two real numbers, \( x_1 \) and \( x_2 \), such that \( f(x_1) = 0 \) and \( f(x_2) = 0 \). It follows from Rolle’s Theorem that there must be a point \( c \) in the interval \([x_1, x_2]\) such that \( f'(c) = 0 \). However,

\[
    f'(x) = 3x^2 + 1, \quad (4.19)
\]

which is always positive, which is a contradiction. Therefore, our assumption that there is more than one real root must be false. This proves that \( f(x) \) has only one real root.

Example 4.13 Prove that a polynomial function of degree \( n \) has at most \( n \) real roots.

Solution We use the technique of mathematical induction, that was illustrated in the Section 3.3. We first prove that a polynomial of degree 0, which is a nonzero constant function, has no real roots. If \( f(x) = c \), where \( c \) is a nonzero constant, then clearly there is no \( x \) for which \( f(x) = 0 \), since \( f(x) = c \) for all \( x \) and \( c \neq 0 \). This is the basis step for the induction.

We now perform the induction step. We assume that the result is true for polynomials of degree \( k \); that is, we assume that a polynomial of degree \( k \) has at most \( k \) real roots. We will now prove that a polynomial of degree \( k + 1 \) has at most \( k + 1 \) real roots.

We proceed as in the previous example, and assume that \( f(x) \) is a polynomial of degree \( k + 1 \) that has more than \( k + 1 \) real roots. Then there are at least \( k + 2 \) real numbers \( x_1, x_2, \ldots, x_{k+2} \), with \( x_1 < x_2 < \cdots < x_{k+2} \), such that \( f(x_1) = f(x_2) = \cdots = f(x_{k+2}) = 0 \). By Rolle’s Theorem, there exists a point \( c \) in each interval \([x_1, x_2], [x_2, x_3], \ldots, [x_{k+1}, x_{k+2}]\) such that \( f'(c) = 0 \). In other words, \( f'(x) \) has at least \( k + 1 \) real roots, since there are \( k + 1 \) intervals. However, \( f'(x) \) is a polynomial of degree \( k \), since \( f(x) \) is a polynomial of degree \( k + 1 \). We have just shown that a polynomial of degree \( k \) has \( k + 1 \) real roots, which contradicts our assumption in the previous paragraph that a polynomial of degree \( k \) has at most \( k \) real roots. Therefore, \( f(x) \), a polynomial of degree \( k + 1 \), has at most \( k + 1 \) real roots.

We have just shown the following:

1. The basis step: The statement we wish to prove, that a polynomial of degree \( n \), is true for \( n = 0 \).

2. The induction step: If statement is true for \( n = k \), then it must be true for \( n = k + 1 \).
By the principle of mathematical induction, the statement must therefore be true for all \( n \). We have, in a sense, created a domino effect. We showed that the first domino falls, and we showed that any domino falling causes the next one to fall. Therefore, all dominos must fall. □

4.3 How Derivatives Affect the Shape of a Graph

In many cases, it is helpful to be able to graph a given function \( f(x) \) correctly, but this can be tedious to accomplish using the straightforward approach of computing several points on the graph of \( f \). In this section, we will see how the first and second derivatives of a twice-differentiable function \( f \) can provide valuable information that can be used to more efficiently obtain a graph that illustrates the behavior of \( f \). As functions are used to model relationships between quantities in some application, it can be extremely beneficial to be able to obtain insight into such relationships efficiently, and informative graphs can provide this insight.

We already know that the first derivative of a function \( f(x) \) can be interpreted as the rate of change of \( y = f(x) \) with respect to \( x \). We have also seen that when \( f'(c) = 0 \) at a given point \( c \), it implies that \( f(x) \) has stopped changing at \( c \). In other words, the function has stopped increasing or decreasing at this particular point. This suggests that when the derivative is nonzero, the function is increasing or decreasing. This is in fact true, as we state precisely in the following result.

**Theorem 4.7** (Increasing/Decreasing Test) Let \( f \) be a function that is differentiable on an interval \((a, b)\).

- If \( f'(x) > 0 \) on \((a, b)\), then \( f \) is increasing on \((a, b)\); that is, if \( x_1 \) and \( x_2 \) are in \((a, b)\) and \( x_1 < x_2 \), then \( f(x_1) < f(x_2) \).
- If \( f'(x) < 0 \) on \((a, b)\), then \( f \) is decreasing on \((a, b)\); that is, if \( x_1 \) and \( x_2 \) are in \((a, b)\) and \( x_1 < x_2 \), then \( f(x_1) > f(x_2) \).

**Proof** Let \( x_1 \) and \( x_2 \) be any two points in \((a, b)\), with \( x_1 < x_2 \). Then, because \( f \) is differentiable on \((a, b)\), then \( f \) is continuous on the closed interval \([x_1, x_2]\) and differentiable on the open interval \((x_1, x_2)\). It follows that we can apply the Mean Value Theorem to conclude

\[
f(x_2) - f(x_1) = f'(c)(x_2 - x_1),
\]

where \( c \) is in \((x_1, x_2)\). If \( f'(x) > 0 \) on \((a, b)\), then certainly \( f'(c) > 0 \), and since \( x_2 > x_1 \), it follows that the right side of the above equation is positive,
and therefore we must have \( f(x_2) - f(x_1) > 0 \), or \( f(x_2) > f(x_1) \). Since this applies to any two points \( x_1 \) and \( x_2 \) in \((a,b)\), we conclude that \( f \) must be increasing on \((a,b)\). If \( f'(x) < 0 \) on \((a,b)\), then we can use similar logic to conclude that \( f \) must be decreasing on \((a,b)\). \( \square \)

Knowledge of when a function \( f(x) \) is increasing or decreasing can be used to determine where \( f \) has a local minimum or local maximum. For instance, a local minimum occurs when \( f \) stops decreasing, and starts increasing. From the previous result, it follows that a local minimum occurs when the derivative changes sign from negative to positive. Naturally, such a sign change can occur when the derivative is equal to zero, but it can also occur when \( f \) fails to be differentiable at a point. In other words, a sign change in the derivative can occur at a critical point \( c \) such that \( f'(c) = 0 \) or \( f'(c) \) does not exist, as we learned in Section 4.1. This characterization of a local minimum or local maximum by sign changes in the derivative is made more precise in the following result, known as the First Derivative Test.

**Theorem 4.8 (First Derivative Test)** Suppose that \( c \) is a critical number of a continuous function \( f \), and that \( f \) is differentiable in an interval containing \( c \), except possibly at \( c \) itself.

- If \( f' \) changes sign from positive to negative at \( c \), then \( f \) has a local maximum at \( c \).
- If \( f' \) changes sign from negative to positive at \( c \), then \( f \) has a local minimum at \( c \).
- If \( f' \) does not change sign at \( c \), then \( f \) does not have a local maximum or local minimum at \( c \).

**Remark** The assumption that \( f \) is continuous at \( c \), but need not be differentiable at \( c \), is very important. This allows for the possibility that \( f \) has a sharp corner in its graph, which would still be a local maximum or minimum. For example, \(|x|\) is differentiable at all \( x \) except at \( x = 0 \), and its derivative changes sign from negative to positive at 0, and it does in fact have a local minimum at that point. \( \square \)

The second derivative of a function can also provide valuable information about the graph. While it is obviously very helpful to know whether a function is increasing or decreasing when attempting to graph it, this knowledge alone is not enough to determine what type of curve should be graphed. It
is also necessary to know whether the rate of increase or decrease is itself increasing or decreasing, as the following example illustrates.

**Example 4.14** In Figure 4.7, the graphs of two increasing functions are shown. In the left graph, \( y = x^3 \) is increasing on the interval \((-3, -1)\) while \( y = x^4 \) is increasing on the interval \((1, 3)\). However, it is clear that the graphs have a different nature, even though both functions are increasing. The difference is illustrated by the relationships between the graphs and their tangent lines. The graph of \( y = x^3 \) lies below the tangent line at the point \((-2, -8)\), while the graph of \( y = x^4 \) lies above the tangent line at the point \((2, 16)\). It can also be observed from the graphs that while \( x^3 \) is increasing, it is increasing more slowly as \( x \) increases, whereas \( x^4 \) is not only increasing, but also increasing at a faster rate as \( x \) increases. □

![Figure 4.7: Graphs of two increasing functions](image)

The observations made in the previous example lead to the following definitions, which aid in the discussion of exactly how a function is increasing or
Definition 4.4 (Concavity) The graph of a function $f$ is said to be **concave upward** on an interval $(a, b)$ if it lies above each of its tangent lines on $(a, b)$. Similarly, we say that the graph of $f$ is **concave downward** on $(a, b)$ if it lies below each of its tangent lines on $(a, b)$.

Example 4.15 As can be seen in Figure 4.7, the graph of the function $y = x^3$ lies below its tangent lines on the interval $(-3, -1)$, and therefore the graph is concave downward. On the other hand, the graph of $y = x^4$ lies above its tangent lines on the interval $(1, 3)$, so the graph is concave upward. \(\square\)

As the previous examples indicate, a function whose graph is concave upward is not only increasing, but is also increasing at a faster rate. The first derivative indicates the rate of increase, so if this rate is increasing, that implies that the derivative of the first derivative, the second derivative, must be positive. Similarly, if the graph of a function is concave downward, its rate of increase is decreasing, and therefore the second derivative must be negative. This leads to the following result.

Theorem 4.9 (Concavity Test) Let $f$ be a function whose second derivative exists on an interval $(a, b)$.

- If $f''(x) > 0$ on $(a, b)$, then the graph of $f$ is concave upward on $(a, b)$.
- If $f''(x) < 0$ on $(a, b)$, then the graph of $f$ is concave downward on $(a, b)$.

Just as we use the critical points of a function can be used to identify possible changes of sign in the first derivative, which imply a change in whether the function is increasing or decreasing, we can use points at which the second derivative changes sign to determine where the graph of a function changes its concavity from upward to downward, or vice versa. We now define these points at which concavity changes.

Definition 4.5 (Inflection Point) A point $(c, f(c))$ on the graph of a function $f$ is called an **inflection point** if $f$ changes concavity at $P$; that is, the graph changes from concave upward to concave downward, or vice versa, at $P$. 

Increasing.
Suppose that a differentiable function \( f \) has a critical point at \( c \); that is, \( f'(c) = 0 \). Then \( f \) has a horizontal tangent line at \( c \). If the graph of \( f \) is concave upward near \( c \), then the graph of \( f \) lies above all of its tangent lines near \( c \). In particular, it lies above the horizontal tangent line at \( c \), which has the equation \( y = f(c) \). Since the graph lies above this line, it follows that \( f(x) \geq f(c) \) for \( x \) near \( c \). In other words, \( f \) has a local minimum at \( c \). This leads to the following result that is useful for characterizing critical points of a function as local minima or local maxima.

**Theorem 4.10** (Second Derivative Test) Let \( f \) be a function whose second derivative exists and is continuous near a point \( c \).

- If \( f'(c) = 0 \) and \( f''(c) > 0 \), then \( f \) has a local minimum at \( c \).
- If \( f'(c) = 0 \) and \( f''(c) < 0 \), then \( f \) has a local maximum at \( c \).

**Remark** If \( f'(c) = 0 \) and \( f''(c) = 0 \), then the Second Derivative Test is inconclusive; \( f \) could have a local minimum, local maximum, or neither. In this case, the First Derivative Test can be used instead. \( \Box \)

**Example 4.16** Figure 4.8 shows the graphs of \( y = x^2 \), \( y = x^3 \), and \( y = x^4 \). All three functions have a critical point at \( x = 0 \), since the first derivative is equal to 0 at that point. The function \( y = x^2 \) has a second derivative of \( y'' = 2 \). Since \( y'' > 0 \), we can conclude from the Second Derivative Test that this function has a local minimum at 0. For \( y = x^3 \), we have \( y'' = 6x \), and so \( y'' = 0 \) at \( x = 0 \). Therefore, the Second Derivative Test is inconclusive. Since \( y = x^4 \) has a second derivative of \( y = 12x^2 \), the Second Derivative Test is inconclusive for this function as well.

For both of these functions, we can instead use the First Derivative Test. For \( y = x^3 \), we have \( y' = 3x^2 \), which is positive except at \( x = 0 \). It follows that the function does not have a local maximum or local minimum at 0, since \( y' \) does not change sign. For \( y = x^4 \), we have \( y' = 4x^3 \), which is positive for \( x > 0 \) and negative for \( x < 0 \). By the First Derivative Test, this function has a local minimum at 0, since \( y' \) changes sign from negative to positive. \( \Box \)

We have seen that the first and second derivatives can provide valuable insight into the behavior of a function on a *bounded* interval \([a, b]\), where the endpoints \( a \) and \( b \) are both finite. This insight greatly facilitates the arduous task of graphing the function on an interval, since it indicates where the function has maxima and minima on the interval, as well as how to draw
Figure 4.8: Graphs of three functions that have a critical point at $x = 0$. The function $y = x^2$ can be shown to have a local minimum at 0 using the Second Derivative Test. The Second Derivative Test is inconclusive for $y = x^3$ and $y = x^4$, and therefore the First Derivative Test must be used to determine whether the critical point corresponds to a local maximum, local minimum, or neither.

the graph in portions of the interval where the function is increasing or decreasing. It follows that a useful, informative graph can be drawn even if the values of the function are known at only a few points. In the next section we will learn how to determine the behavior of a function on an unbounded interval such as the entire real number line.

Example 4.17 Consider the function $f(x) = x^2$. Its derivative is $f'(x) = 2x$, which is positive for $x > 0$, negative for $x < 0$, and equal to 0 when $x = 0$. By the First Derivative Test, $f$ has a local minimum at $x = 0$, since $f'$ changes sign from negative to positive at $x = 0$.

On the other hand, $f(x) = |x|$ is not differentiable at $x = 0$. For $x < 0$,
f'(x) = -1, and for x > 0, f'(x) = 1. Since f' changes sign from negative to positive at x = 0, the First Derivative Test can still be used to conclude that it has a local minimum at x = 0.

Both functions are shown in Figure 4.9. □

Figure 4.9: The graphs of \( f(x) = x^2 \) and \( f(x) = |x| \). Both functions have a local minimum at \( x = 0 \), since in each case the first derivative changes sign from negative to positive. The fact that \( x^2 \) is differentiable at \( x = 0 \), while \( |x| \) is not, is irrelevant.

Example 4.18 The functions \( f(x) = \sin x \) and \( g(x) = \cos x \) are shown in Figure 4.10. The Increasing/Decreasing Test states the following:

- If \( f'(x) > 0 \) on an interval, then \( f \) is increasing on that interval, and
- If \( f'(x) < 0 \) on an interval, then \( f \) is decreasing on that interval.

We have \( g'(x) = -\sin x \). Since \( \sin x > 0 \) on the interval \((0, \pi)\), we have \( g'(x) < 0 \) on \((0, \pi)\), and therefore \( g(x) = \cos x \) is decreasing on \((0, \pi)\), by the
CHAPTER 4. APPLICATIONS OF DERIVATIVES

Figure 4.10: The graphs of $\sin x$ (solid curve) and $\cos x$ (dotted curve).

Increasing/Decreasing Test. This can be confirmed by examining the figure. Similarly, $f'(x) = \cos x$, which is positive on the interval $[0, \pi/2)$ and negative on the interval $(\pi/2, 3\pi/2)$. Therefore, by the Increasing/Decreasing Test, $f$ is increasing on $[0, \pi/2)$ and decreasing on $(\pi/2, 3\pi/2)$, as can be confirmed from the figure. □

4.4 A Summary of Curve Sketching

In previous sections we have learned how to use limits and derivatives (which are limits themselves) to gain insight into the behavior of the graph of a given function $f(x)$ on a given domain, whether that domain is bounded or unbounded. In this section we will now summarize all of these techniques for understanding a function’s graph and describe a systematic procedure one can follow in order to simply obtain an accurate, informative graph. Since a function is typically used to model a relationship between quantities in
some application, such knowledge of a function’s graph can provide valuable insight into such a relationship, and by extension, greater understanding of the underlying application.

It is tempting to think that a useful graph of a function \( f(x) \) can be obtained simply by evaluating \( f(x) \) at several \( x \)-values, plotting the resulting points, and then connecting them with a smooth curve. However, it is possible that important behavior of the function can be missed by such an approach. For example, what if a function has a vertical asymptote at \( x = a \), where \( a \) is not among the \( x \)-values at which \( f \) is evaluated to produce the graph? Then the graph is misleading. Using calculus, however, all “interesting” behavior of a function can be understood and incorporated into a graph.

In graphing a given function \( f(x) \), one should ask the following questions:

- **What is the domain of the function?** It is important to know where the function is defined, and not defined. If, for example, \( f(x) \) is only defined for \( x \geq 0 \), then we can ignore negative values of \( x \) in constructing our graph.

- **What are the intercepts of the function?** Often a graph will include the \( x \)-axis and \( y \)-axis in order to provide a well-known frame of reference. Therefore, it is natural to indicate the points at which the graph of \( f(x) \) intersects these axes, if they can easily be determined.

- **Is the graph symmetric in any sense?** Any symmetry in the graph of \( f(x) \) can reduce the effort of graphing significantly, as we can simply construct a portion of the graph and use this symmetry to easily obtain the remainder. For example, if \( f \) is an even function, we have \( f(-x) = f(x) \), so we can simply graph the function for positive \( x \) and then reflect it across the vertical line \( x = 0 \) to obtain the graph for negative \( x \).

Another type of symmetry is *periodicity*, in which \( f(x) = f(x + L) \) for all \( x \) in its domain and some fixed value \( L \), which is called the period of \( f \). The graph of such a function \( f \) can be obtained from the graph of \( f \) on an interval of length \( L \), which is then repeated in adjoining intervals of length \( L \) throughout the domain of \( f \).

- **Does the function have any asymptotes?** Certainly knowledge of how the function behaves as \( x \) becomes infinite can be useful in constructing its graph on an unbounded domain, whether there is a horizontal asymptote or the function itself becomes infinite. Similarly, it is desirable to know if there exist any points at which the function value
becomes infinite; that is, it has a vertical asymptote. As mentioned previously, such important behavior can easily be missed when constructing a graph by simply plotting points. When \( f(x) \) has a vertical asymptote at \( x = a \), it is important to determine whether \( f(x) \) approaches \( \infty \) or \( -\infty \) as \( x \) approaches \( a \), from either direction, in order to obtain a correct graph.

- **When is the function increasing or decreasing?** This can be answered by computing \( f'(x) \) and determining where it is positive or negative. This determines the general direction of the curve \( y = f(x) \) throughout the domain.

- **Where does the function have a local minimum or maximum?** Once it is determined where the function is increasing or decreasing, points at which the function changes from increasing to decreasing indicate the location of local maximma, while changes from decreasing to increasing indicate the location of local minima.

In fact, it is wise to compute all of the critical numbers of \( f(x) \), since they include all points at which local maxima or minima occur. In addition, if it is not easy to determine algebraically where \( f \) is increasing or decreasing (i.e., where \( f'(x) > 0 \) and \( f'(x) < 0 \)), then one can instead compute the critical numbers, and then evaluate \( f' \) at points near those numbers to determine where \( f \) is increasing or decreasing.

- **What is the concavity of the graph?** Determining where \( f \) is increasing or decreasing determines the general direction of the curve \( y = f(x) \), but it does not indicate the “shape” or “curvature” of the graph. This can be determined from the concavity of the graph, which is indicated by the sign of the second derivative.

By obtaining this information about a given function, we can easily construct a graph that captures all of its “interesting” behavior, as we illustrate with an example.

**Example 4.19** Consider the function

\[
 f(x) = \frac{1}{x^2 - 4}. \quad (4.21)
\]

The domain of this function is the entire real number line, except for \( x = \pm 2 \).

The \( y \)-intercept of \( f(x) \) can be obtained by substituting \( x = 0 \), which yields \( f(0) = -1/4 \). The \( x \)-intercept can be obtained by determining where
4.4. A SUMMARY OF CURVE SKETCHING

the numerator is equal to zero; this does not occur at any \( x \), so there are no \( x \)-intercepts.

Due to the squaring of \( x \) in the denominator, we can see that \( f \) is an \textit{even} function; that is, \( f(x) = f(-x) \). Therefore the graph is symmetric with respect to the vertical line \( y = 0 \).

By computing the limits of \( f(x) \) as \( x \) approaches \( \pm 2 \) from the left and right, we can determine that \( f \) has vertical asymptotes at these points. Specifically,

\[
\lim_{x \to -2^-} f(x) = \infty, \quad \lim_{x \to -2^+} f(x) = -\infty, \quad (4.22)
\]

\[
\lim_{x \to 2^-} f(x) = -\infty, \quad \lim_{x \to 2^+} f(x) = \infty. \quad (4.23)
\]

As \( x \) approaches \( \infty \), the denominator becomes larger and larger, while the numerator remains fixed at 1. It follows that

\[
\lim_{x \to \infty} f(x) = 0. \quad (4.24)
\]

Since \( f \) is an even function, we can also conclude that

\[
\lim_{x \to -\infty} f(x) = 0. \quad (4.25)
\]

In other words, \( f \) has one horizontal asymptote, at \( y = 0 \).

To determine where \( f(x) \) is increasing and decreasing, and if it has any local maxima or minima, we compute its derivative, which yields

\[
f'(x) = -\frac{2x}{(x^2 - 4)^2}. \quad (4.26)
\]

We can easily determine that \( f'(x) = 0 \) at \( x = 0 \). This is the only critical number of \( f \), and therefore the only point at which a local maximum or minimum can occur. It can be determined by evaluating \( f'(x) \) at points near zero and the asymptotes at \( x = \pm 2 \) that \( f \) is increasing on the intervals \(( -\infty, -2) \) and \((-2, 0)\), and decreasing on the intervals \((0, 2)\) and \((2, \infty)\).

Since \( f'(x) \) changes sign from positive to negative at \( x = 0 \), we can conclude that \( f \) has a local maximum at \( 0 \), with maximum value \( f(0) = -1/4 \).

If we compute the second derivative, we obtain

\[
f''(x) = -\frac{(x^2 - 4)^2(2) - (2x)(2)(x^2 - 4)(2x)}{(x^2 - 4)^4} = -\frac{6x^4 + 16x^2 + 32}{(x^2 - 4)}. \quad (4.27)
\]
CHAPTER 4. APPLICATIONS OF DERIVATIVES

Using the quadratic formula, we can determine where the numerator \(-6x^4 + 16x^2 + 32\) is equal to 0. We have

\[
x^2 = \frac{-16 \pm \sqrt{16^2 - 4(-6)(32)}}{-12} = \frac{-16 \pm \sqrt{256 + 768}}{-12} = \frac{-16 \pm 32}{-12}.
\] (4.28)

Choosing the minus sign yields \(x^2 = 4\), but this is satisfied when the denominator of \(f''(x)\) is equal to zero. Evaluating \(f''(x)\) at points on either side of each asymptote, we can determine that the graph of \(f\) is concave upward when \(x < 2\) and \(x > 2\), and concave downward when \(-2 < x < 2\).

Using all of the information we have gathered, we can now construct a graph of \(f(x)\). This graph is shown in Figure 4.11.

![Figure 4.11: Graph of \(y = 1/(x^2 - 4)\). The dashed lines indicate asymptotes. The local maximum at \(x = 0\) is indicated by a circle.](image)

**Example 4.20** Graph the function

\[f(x) = x^4 - 8x^3 + 22x^2 - 24x + 10.\] (4.29)
Solution We examine several aspects of this function before drawing the graph:

- **Intercepts:** It can be shown that \( f(x) \) has no \( x \)-intercepts; that is, the graph does not cross the \( x \)-axis. The \( y \)-intercept can be found by computing \( f(0) = 10 \).

- **Symmetry:** There is no obvious symmetry in this function; by computing \( f(-x) = x^4 + 8x^3 + 22x^2 + 24x + 10 \), it can be seen that \( f(x) \) is neither an even function nor an odd function.

- **Asymptotes:** Since \( f(x) \) is a polynomial, it is defined on the entire real number line, so it does not have any vertical asymptotes. It does not have any horizontal asymptotes either, since

\[
\lim_{x \to \infty} f(x) = \infty, \quad \text{and} \quad \lim_{x \to -\infty} f(x) = \infty. \tag{4.30}
\]

- **First Derivative:** The first derivative is

\[
f'(x) = 4x^3 - 24x^2 + 44x - 24 = 4(x^3 - 6x^2 + 11x - 6). \tag{4.31}\]

We need to determine the critical numbers of \( f \), which are the points at which \( f'(x) = 0 \) or \( f'(x) \) does not exist. Since \( f \) is a polynomial, it is differentiable for all \( x \), so we only need to find the points at which \( f'(x) = 0 \). This is the case when \( x = 1, x = 2 \) and \( x = 3 \). By evaluating \( f'(x) \) at other points, we can determine that \( f'(x) < 0 \) for \( x < 1 \) and \( 2 < x < 3 \), and \( f'(x) > 0 \) for \( 1 < x < 2 \) and \( x > 3 \). In other words, \( f(x) \) is decreasing on \((-\infty, 1)\) and \((2, 3)\), while it is increasing on \((1, 2)\) and \((3, \infty)\).

Since \( f'(x) \) changes sign from negative to positive at \( x = 1 \), and at \( x = 3 \), it follows from the First Derivative Test that \( f(x) \) has local minima at these two points. The same test can be used to show that \( f(x) \) has a local maximum at \( x = 2 \). Evaluating \( f(x) \) at these points, we obtain \( f(1) = 1, f(2) = 2 \) and \( f(3) = 1 \).

- **Second Derivative:** The second derivative is

\[
f''(x) = 4(3x^2 - 12x + 11). \tag{4.32}\]

Using the quadratic formula, it can be shown that \( f''(x) = 0 \) when \( x = 2 \pm \sqrt{3}/3 \), or \( x \approx 1.4226 \) and \( x \approx 2.5774 \). Computing \( f''(1) = 8 \), \( f''(2) = -4 \) and \( f''(3) = 8 \) indicates that \( f''(x) > 0 \) for \( x < 1.4226 \) and
$x > 2.5774$, while $f''(x) < 0$ for $1.4226 < x < 2.5774$. In other words, the graph of $f(x)$ is concave upward on $(-\infty, 1.4226)$ and $(2.5774, \infty)$ and concave downward on $(1.4226, 2.5774)$. The points at which $f''(x) = 0$ are inflection points, since the concavity of the graph changes at these points. The values of $f''(x)$ at $x = 1$, $x = 2$ and $x = 3$ can be used to confirm the local minima at $x = 1$ and $x = 3$ and the local maximum at $x = 2$, by the Second Derivative Test.

The resulting graph is shown in Figure 4.12.

![Figure 4.12: Graph of $y = x^4 - 8x^3 + 22x^2 - 24x + 10$](image)

**Example 4.21** Graph the function

$$f(x) = \frac{x + 2}{x + 1}. \quad (4.33)$$

**Solution** We examine several aspects of this function before drawing the graph:
4.4. A SUMMARY OF CURVE SKETCHING

- **Intercepts:** An x-intercept occurs when \( x = -2 \), since \( f(-2) = 0 \). The y-intercept occurs at \( y = 2 \), since \( f(0) = 2 \).

- **Symmetry:** This function is neither even nor odd, since \( f(-x) = (2 - x)/(1 - x) \) is not equal to \( f(x) \) or \(-f(x)\).

- **Asymptotes:** A vertical asymptote occurs at \( x = -1 \). This can be seen by first noting that \( f(x) \) is undefined at \( x = -1 \), and then computing

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

and

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

These limits can be computed by noting that when \( x \) is near \(-1\), the numerator is positive, but when \( x \) approaches \(-1\) from the left, \( x < -1 \), so \( x + 1 < 0 \). Similarly, when \( x \) is approaching \(-1\) from the right, \( x > -1 \), so \( x + 1 > 0 \).

To determine whether there is a horizontal asymptote, we compute

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

Similarly,

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

Either of these limits is sufficient to conclude that \( f(x) \) has a horizontal asymptote at \( y = 1 \), but both limits are needed to determine which portions of the graph of \( f \) are actually approaching this line.

- **First Derivative:** We have, by the Quotient Rule,

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

This derivative is defined wherever \( f(x) \) is defined, and wherever it is defined, it is negative, so \( f(x) \) is decreasing on its entire domain and it has no local maxima or minima.
• **Second Derivative:** We have

\[ f''(x) = \frac{d}{dx}[-(x + 1)^{-2}] = -(-2)(x + 1)^{-3} = \frac{2}{(x + 1)^3}. \]  

(4.38)

For \( x > -1 \), we have \( x + 1 > 0 \), so \( f''(x) > 0 \), while \( f''(x) < 0 \) for \( x < -1 \). In other words, the graph of \( f \) is concave upward on \((-1, \infty)\) and concave downward on \((-\infty, -1)\).

The resulting graph is shown in Figure 4.13.

---

**Example 4.22** Graph the function

\[ f(x) = \sec x \]  

(4.39)

on the interval \([-\pi, \pi]\).
Solution We examine several aspects of this function before drawing the graph:

- **Intercepts:** The $y$-intercept is given by $f(0) = \sec 0 = 1/\cos 0 = 1$. From the fact that $\sec x = 1/\cos x$, we can see that $f(x) \neq 0$ anywhere on its domain, so there are no $x$-intercepts.

- **Symmetry:** Since $\cos x$ is an even function, $\sec x = 1/\cos x$ is also an even function, so the graph of $f$ is symmetric with respect to the $y$-axis.

- **Asymptotes:** On the interval $[-\pi, \pi]$, the only points at which $\sec x$ is undefined are the points at which its denominator, $\cos x$, is equal to 0. These points are $x = \pi/2$ and $x = -\pi/2$. By noting the sign of $\cos x$ as $x$ approaches $\pi/2$ and $-\pi/2$ from the left and right, we obtain the following limits:

  \[
  \lim_{x \to -\pi/2^-} \sec x = -\infty, \quad \lim_{x \to -\pi/2^+} \sec x = \infty, \quad \quad (4.40)
  \]

  \[
  \lim_{x \to \pi/2^-} \sec x = \infty, \quad \lim_{x \to \pi/2^+} \sec x = -\infty. \quad \quad (4.41)
  \]

  We conclude that $f(x)$ has vertical asymptotes at $x = -\pi/2$ and $x = \pi/2$.

  Since we are only concerned with the behavior of $f(x)$ on the interval $[-\pi, \pi]$, we do not need to determine horizontal asymptotes.

- **First Derivative:** We have

  \[ f'(x) = \sec x \tan x. \quad \quad (4.42) \]

  Since $\sec x$ is never equal to zero, $f'(x) = 0$ when $\tan x = 0$, which is only the case if $\sin x = 0$, since $\tan x = \sin x/\cos x$. This is the case when $x = 0$, $x = \pi$, and $x = -\pi$. Evaluating $f'(x)$ at points that are near these critical numbers, we can determine that $f'(x) > 0$ when $0 < x < \pi/2$ and $\pi/2 < x < \pi$, while $f'(x) < 0$ when $-\pi < x < -\pi/2$ and $-\pi/2 < x < 0$. In other words, $f(x)$ is increasing on $(0, \pi/2)$ and $(\pi/2, \pi)$ and decreasing on $(-\pi, -\pi/2)$ and $(-\pi/2, 0)$. We will use the second derivative to determine the significance of the critical numbers of $f$. 
• **Second Derivative:** Using the Product Rule, we obtain

\[
f''(x) = \frac{d}{dx} [\sec x \tan x]
\]

\[
= \tan x \frac{d}{dx} [\sec x] + \sec x \frac{d}{dx} [\tan x]
\]

\[
= \tan x (\sec x \tan x) + \sec x \sec^2 x
\]

\[
= \sec x (\tan^2 x + \sec^2 x).
\]

Since \(\sec x \neq 0\) and \(\tan^2 x + \sec^2 x > 0\) for all \(x\), \(f''(x)\) is never equal to zero. Also, the sign of \(f''(x)\) is determined by the sign of the \(\sec x\) factor, which in turn is determined by the sign of \(\cos x\).

We will use this fact to apply the Second Derivative Test to the critical points \(x = -\pi, 0, \pi\) that were found above. At \(x = -\pi\) and \(x = \pi\), \(\cos x = -1\), so \(f''(x) < 0\). It follows that \(f(x)\) has local maxima at \(x = -\pi\) and \(x = \pi\), and has a graph that is concave downward on \([-\pi, -\pi/2)\) and \((\pi/2, \pi]\). At \(x = 0\), \(\cos x = 1\), so \(f''(x) > 0\), which implies that \(f(x)\) has a local minimum at \(x = 0\). Also, we can conclude that the graph of \(f\) is concave upward on \((-\pi/2, \pi/2)\).

The resulting graph is shown in Figure 4.14. □

### 4.5 Optimization

In Section 4.1, we developed a step-by-step procedure for finding the absolute maximum and absolute minimum of a function \(f(x)\) on a given interval \([a, b]\). This procedure is a key ingredient in the solution of a very important problem called the optimization problem. The optimization problem is to find the values of a set of variables \(x_1, x_2, \ldots, x_n\) such that the objective function

\[
f(x_1, x_2, \ldots, x_n)
\]

attains its minimum or maximum value, subject to constraints on the variables \(x_1, \ldots, x_n\).

In general, this is a very difficult problem to solve, and there is much ongoing research devoted to the development of more effective solution techniques. However, most such techniques are based on the simple procedure that is used to find the maximum or minimum value of a function of one variable, \(f(x)\), on an interval \([a, b]\).

In this section, we will focus on a simplified version of the general optimization problem, in which it is possible to express the objective function
4.5. OPTIMIZATION

Figure 4.14: Graph of \( y = \sec x \) on the interval \([-\pi, \pi]\). The dashed lines indicate vertical asymptotes.

\[ f(x_1, x_2, \ldots, x_n) \] as a function of a single variable \( x \), and then find the maximum or minimum value of this function on an interval. We illustrate this process with an example.

**Example 4.23** Suppose that a fence of length \( L \) is to be constructed, and that it will enclose a rectangular area. Prove that the fence will enclose the largest area possible if the enclosed area is actually a perfect square.

**Solution** In this problem, we wish to find the values of the variables \( \ell \) and \( w \), representing the length and width of the region enclosed by the fence, so that the area \( A \) of the region, defined by the function

\[ A = f(\ell, w) = \ell w, \quad (4.44) \]

attains its maximum value.
The function $f$ depends on two variables, but we only know how to work with functions of a single variable. We can eliminate one variable by taking into account the fact that $\ell$ and $w$ are both subject to the constraint that the perimeter of the fence, given by $2\ell + 2w$, is equal to $L$. Therefore, we have

$$\ell = \frac{L - 2w}{2} = \frac{L}{2} - w.$$  \hspace{1cm} (4.45)

We can then express our objective function as a function of one variable, $w$, as follows:

$$f(w) = \left(\frac{L}{2} - w\right)w = \frac{L}{2}w - w^2.$$  \hspace{1cm} (4.46)

We know how to find the maximum value of this function on any given interval, but we need to determine the correct interval in this case. Certainly, $w$ cannot be negative, since it represents a width. Furthermore, $w$ cannot exceed $L/2$, since the width accounts for two sides of the rectangular region and the perimeter must be equal to $L$. Therefore, we will determine the maximum value of $f(w)$ on the interval $[0, L/2]$.

Using the procedure introduced in Section 4.1, we find all critical points of $f(w)$ on this interval. We have

$$f'(w) = \frac{d}{dw} \left[ \frac{L}{2}w - w^2 \right] = \frac{L}{2} - 2w,$$  \hspace{1cm} (4.47)

and find that $f'(w) = 0$ when $L/2 = 2w$, or $w = L/4$. Evaluating $f(w)$ at this critical point yields

$$A = f\left(\frac{L}{4}\right) = \frac{L}{2} \cdot \frac{L}{4} - \left(\frac{L}{4}\right)^2 = \frac{L^2}{8} - \frac{L^2}{16} = \frac{L^2}{16}.$$  \hspace{1cm} (4.48)

The only other points on $[0, L/2]$ at which $f(w)$ can have an absolute maximum are the endpoints $w = 0$ and $w = L/2$, but since $f(0) = 0$ and $f(L/2) = 0$, it follows that $A$ is maximized when $w = L/4$, with maximum value $A = f(L/4) = L^2/16$. From the relation $\ell = L/2 - w$, we obtain $\ell = L/2 - L/4 = L/4$, so the width and the height of the rectangular region of maximum area are equal. In other words, the region is a square. \Box

The preceding example suggests a general procedure for solving optimization problems. Note the similarities between the following procedure and the procedure given in Section 3.9 for solving related-rates problems.

1. **Understand the problem.** What is the quantity that is to be optimized? (that is, what needs to be maximized or minimized?) How does it
depend on other quantities? Are there constraints on the values of any quantities? How do all of the quantities in the problem relate to one another?

In the preceding example, the quantity to be maximized was area of the region enclosed by the fence, and the other given quantity was the perimeter of the fence. Since the region was assumed to be rectangular, this implied two other relevant quantities: the width and height of the region. The area was known to be the product of the width and height, while the perimeter was known to be the sum of twice the width and twice the height.

2. *Introduce notation.* Assign variable names to all quantities in the problem, for use in subsequent equations. For each variable, determine the range of values that the variable can assume, based on information contained in the problem statement. For instance, if a quantity represents a physical dimension such as length or area, then it cannot be negative. Once variable names have been assigned to the quantities in the problem, any relations among these quantities should then be described mathematically using equations.

In the preceding example, we denoted the length of the region enclosed by the fence by the variable \( \ell \), while the width was denoted by \( w \). The perimeter had a fixed value, \( L \), which was a constant rather than a variable. The area of the region was denoted by the variable \( A \). The relation between the area and the width and height was described by the equation \( A = \ell w \), while the relation between the perimeter and the width and height was described by the equation \( 2\ell + 2w = L \).

3. *Express the objective function in terms of symbols.* Now that the problem statement has been “translated” into the language of mathematics, use the equations obtained in the previous step to obtain a formula for the objective function that clearly indicates how the quantity that is to be optimized depends on other quantities in the problem. In the preceding example, the quantity we wish to maximize was the area \( A \) of the region enclosed by the fence, so the objective function was defined by the equation \( A = \ell w \). Specifically, we defined the objective function \( f(\ell, w) \) by \( f(\ell, w) = \ell w \).

4. *Eliminate all but one variable.* Since we only know how to find the maximum or minimum of a function of one variable, we must use any known relations between the independent variables of the objective
function in order to eliminate all but one of them. Often, constraints on the independent variables can be used to accomplish this. In the preceding example we used the constraint on the perimeter of the region enclosed by the fence in order to express its length in terms of its width, allowing us to eliminate the length from the definition of the objective function that represented the area.

5. Find the absolute maximum or minimum. Having obtained an objective function that depends on only one variable, and knowing the range of values that this variable can assume, we can use the procedure described in Section 4.1 for finding the maximum or minimum value of a function on an interval. We used this procedure in the preceding example to compute the maximum value of the objective function \( f(w) \) that was obtained after eliminating any dependence on the length \( \ell \).

We now illustrate this procedure with another example.

**Example 4.24** Suppose that a cylindrical container, with a closed bottom but open top, is to be made from 9000 cm\(^2\) of material. What height and radius should the cylinder have in order to maximize the container’s volume?

**Solution** The quantity to be optimized is the volume of the cylinder. The cylinder has a height and a radius, which together determine the volume. The height and radius are both constrained by the fact that only a fixed amount of material is available.

We now introduce some notation. Let \( V \) be the volume of the cylinder, \( h \) be the height, and \( r \) be the radius. Certainly, all of these quantities must be positive. To determine whether there are any other constraints on these variables, we examine the condition that the amount of material available is 9000 cm\(^2\). Since the surface area \( S \) of a cylinder of radius \( r \) and height \( h \), and open top, is given by \( S = 2\pi rh + \pi r^2 \), it follows that \( 2\pi rh + \pi r^2 = 9000 \) defines another constraint on the values of \( r \) and \( h \).

The objective function is given by the formula for the volume of a cylinder, \( V = \pi r^2 h \). Our objective function is therefore a function of two variables,

\[
V = f(r, h) = \pi r^2 h. \quad (4.49)
\]

We now rewrite this function so that it depends on only one variable. From the constraint \( 2\pi rh + \pi r^2 = 9000 \), we have

\[
h = \frac{4500}{\pi r} - \frac{r}{2}. \quad (4.50)
\]
4.5. OPTIMIZATION

Substituting this relation into our objective function yields

\[ V = f(r) = \pi r^2 \left( \frac{4500}{\pi r} - \frac{r}{2} \right) = 4500r - \frac{\pi r^3}{2}. \]  \hspace{1cm} (4.51)

Now, we are ready to find the maximum value of \( f(r) \). We have

\[ f'(r) = 4500 - \frac{3\pi}{2} r^2, \]  \hspace{1cm} (4.52)

which is equal to 0 when

\[ r = \pm \sqrt{\frac{4500}{3\pi}} = \pm \sqrt{\frac{3000}{\pi}}. \]  \hspace{1cm} (4.53)

Certainly, we must choose the positive square root to be our potential absolute maximum, since we can only consider positive values of \( r \).

To determine whether this critical number \( r = \sqrt{\frac{3000}{\pi}} \) actually corresponds to a local maximum, we use the Second Derivative Test. We have

\[ f''(r) = -3\pi r, \]  \hspace{1cm} (4.54)

which is negative for \( r > 0 \), so \( f(r) \) has a local maximum at \( r = \sqrt{\frac{3000}{\pi}} \).

Is this local maximum the absolute maximum of \( f(r) \) on the interval \( 0 < r < \infty \)? To answer this, we note that for \( 0 < r < \sqrt{\frac{3000}{\pi}} \), \( f'(r) > 0 \), so \( f \) is increasing, while for \( r > \sqrt{\frac{3000}{\pi}} \), \( f'(r) < 0 \), so \( f \) is decreasing. Therefore, it is not possible for \( f \) to assume a larger value at any other point.

To prove this, we first note note that by the definition of a local maximum given in Section 4.1, \( f(\sqrt{\frac{3000}{\pi}}) \geq f(r) \) for all \( r \) in an open interval containing \( \sqrt{\frac{3000}{\pi}} \). Then, we can use the definition of an increasing function, given in Section 4.3, to conclude that \( f(r) < f(r_0) \) whenever \( 0 < r < r_0 < \sqrt{\frac{3000}{\pi}} \), since \( f(r) \) is increasing on the interval \( 0 < r < \sqrt{\frac{3000}{\pi}} \). Because we can choose the point \( r_0 \) to be arbitrarily close to \( \sqrt{\frac{3000}{\pi}} \), and \( f \) has a local maximum at \( \sqrt{\frac{3000}{\pi}} \), we can always choose \( r_0 \) so that \( f(r_0) \leq f(\sqrt{\frac{3000}{\pi}}) \), which guarantees that \( f(r) \leq f(\sqrt{\frac{3000}{\pi}}) \) for all \( r \) such that \( 0 < r < \sqrt{\frac{3000}{\pi}} \). Similarly, we can use the definition of a decreasing function to conclude that \( f(r) \leq f(\sqrt{\frac{3000}{\pi}}) \) for \( r > \sqrt{\frac{3000}{\pi}} \). Therefore the maximum value of \( f(r) \) on the interval \( 0 < r < \infty \) is attained at \( r = \sqrt{\frac{3000}{\pi}} \).

We conclude that the dimensions of the cylinder with maximum volume are

\[ r = \sqrt{\frac{3000}{\pi}} \approx 30.9019 \text{ cm}, \quad h = \frac{4500}{\pi r} - \frac{r}{2} \approx 30.9019 \text{ cm}. \]  \hspace{1cm} (4.55)

In other words, the height and radius must be equal. \( \square \)
When the domain of the objective function \( f(x) \) is an open interval \((a,b)\), as in the preceding example, it can be difficult to check whether any critical number of \( f \) on \((a,b)\) corresponds to the absolute maximum or minimum of \( f \) on \((a,b)\), since there are no endpoints that can be checked. However, as seen in the preceding example, if it is known that \( c \) is a critical number of \( f \) in \((a,b)\), and that \( f(x) \) is increasing for \( x < c \) and decreasing for \( x > c \), then it is not possible for the absolute maximum of \( f \) on \((a,b)\) to occur at any other point. A similar statement can be made concerning the absolute minimum. This leads to the following result, which can be used to determine the absolute maximum or minimum of a function on an open interval.

**Theorem 4.11 (First Derivative Test for Absolute Extreme Value)** Let \( f \) be continuous on \((a,b)\) and let \( c \) be between \( a \) and \( b \).

- If \( f(x) \) is increasing for \( a < x < c \) and decreasing for \( c < x < b \), then the absolute maximum of \( f \) on \((a,b)\) occurs at \( c \).
- If \( f(x) \) is decreasing for \( a < x < c \) and increasing for \( c < x < b \), then the absolute minimum of \( f \) on \((a,b)\) occurs at \( c \).

We do not present the proof here, but the main ideas of the proof can be found in the preceding example.

**Example 4.25** Find two positive numbers whose product is 100 and whose sum is as small as possible.

**Solution** We proceed by following the general procedure for solving optimization problems.

1. **What is the objective function?** We wish to minimize the sum of the two numbers, which we will denote by \( x \) and \( y \). Therefore, our objective function is
   \[
   f(x, y) = x + y. \quad \text{(4.56)}
   \]

2. **What are the constraints on the variables?** We must have \( x > 0 \) and \( y > 0 \). Also, since the product of \( x \) and \( y \) must equal 100, we have the constraint
   \[
   xy = 100. \quad \text{(4.57)}
   \]

3. **Eliminate variables so that the objective function depends on only one variable.** From the constraint \( xy = 100 \), we have \( y = 100/x \), so we can rewrite our objective function as a function of \( x \):
   \[
   f(x) = x + \frac{100}{x}. \quad \text{(4.58)}
   \]
4. **Find the minimum value of the objective function.** We determine the minimum value of \( f(x) \) on the open interval \((0, \infty)\). To find the absolute minimum, we use the First Derivative Test for Absolute Extreme Value, which is given in Section 4.6. This requires finding a point \( c \) in the interval \((0, \infty)\) such that \( f(x) \) is decreasing for \( x < c \) and \( f(x) \) is increasing for \( x > c \). This would be the case if \( f'(x) < 0 \) for \( x < c \), and \( f'(x) > 0 \) for \( x > c \). Therefore, we first find a point at which \( f'(x) = 0 \) and then determine how \( f'(x) \) changes sign.

We have
\[
f'(x) = \frac{d}{dx} \left[ x + \frac{100}{x} \right] = 1 - \frac{100}{x^2}.
\] (4.59)

From the equation
\[
1 - \frac{100}{x^2} = 0,
\] (4.60)
we obtain
\[
x^2 = 100,
\] (4.61)
or \( x = 10 \). This is the only critical number of \( f(x) \) on the interval \((0, \infty)\).

To confirm that this point corresponds to the minimum value of \( f \) on \((0, \infty)\), there are two ways we can proceed:

- **We can use the First Derivative Test.** On the interval \((0, \infty)\), \( f'(x) = 0 \) only at \( x = 10 \). It is therefore nonzero everywhere else on \((0, \infty)\); we just need to determine what sign it is. Since \( f'(x) \) cannot change sign on \((0, 10)\), we can evaluate \( f'(x) \) at any point in \((0, 10)\), such as \( x = 1 \), to determine whether \( f' \) is positive or negative on that interval. Since \( f'(1) = 1 - 100/1 = -99 \), we conclude that \( f'(x) < 0 \) on \((0, 10)\); that is, \( f(x) \) is decreasing for \( 0 < x < 10 \). Similarly, we can show that \( f'(x) > 0 \) on \((10, \infty)\); that is, \( f(x) \) is increasing for \( x > 10 \).

- **We can use the Second Derivative Test.** We have
\[
f''(x) = \frac{d}{dx} \left[ 1 - \frac{100}{x^2} \right] = \frac{d}{dx}[1-100x^{-2}] = -2(-100)x^{-3} = \frac{200}{x^3},
\] (4.62)
which is positive on \((0, \infty)\). Therefore, \( f(x) \) has a local minimum at \( x = 10 \). Furthermore, because \( f''(x) > 0 \) on \((0, \infty)\), \( f'(x) \) is increasing, which implies that at \( x = 10 \), not only is \( f'(x) = 0 \), but \( f' \) is changing sign from negative to positive. That is, \( f(x) \) is decreasing for \( 0 < x < 10 \) and increasing for \( x > 10 \).
In either case, we reach the conclusion that $f(x)$ is decreasing for $0 < x < 10$ and increasing for $x > 10$. Therefore, by the First Derivative Test for Absolute Extreme Value, the absolute minimum of $f(x)$ on $(0, \infty)$ occurs at $x = 10$.

From the relation $xy = 100$, we obtain $y = 100/x = 100/10 = 10$, so the two positive numbers whose product is 100 and whose sum is minimum are $x = 10$ and $y = 10$. □

**Example 4.26** Suppose that 1000 m of fencing is used to enclose a rectangular region of width $w$ and length $l$, with the area bisected by a segment that is also of length $l$ (see Figure 4.15). How should the dimensions of the rectangular region be chosen so that it encloses the largest possible area?

![Figure 4.15](image)

Figure 4.15: The rectangular region shown is to be enclosed by a fence that is 1000 m long, including the portion that bisects the region.

**Solution** We proceed by following the general procedure for solving optimization problems.
1. **What is the objective function?** We wish to maximize the area $A$ of the rectangular region, which has length $\ell$ and width $w$, so our objective function is

$$A(\ell, w) = \ell w.$$  \hfill (4.63)

2. **What are the constraints on the variables?** There are only 1000 m of fencing available, so the perimeter of the region, as well as the length of the segment that bisects the region, must equal 1000. In other words,

$$2w + 3\ell = 1000.$$ \hfill (4.64)

Certainly, $w$ and $\ell$ must both be positive.

3. **Eliminate variables until so that the objective function depends on only one variable.** Solving for $w$ in the above constraint yields

$$w = \frac{1000 - 3\ell}{2}.$$ \hfill (4.65)

This allows us to rewrite our objective function as a function of $\ell$,

$$f(\ell) = \ell w = \ell \frac{1000 - 3\ell}{2} = \frac{1}{2}(1000\ell - 3\ell^2).$$ \hfill (4.66)

4. **Find the maximum value of the objective function.** We now seek the absolute maximum of $f(\ell)$ on the interval $0 < \ell < \infty$. As in the previous example, we will use the First Derivative Test for Absolute Extrema, which requires us to determine where $f$ is increasing and decreasing. To accomplish this, we find all critical numbers of $f(\ell)$ on $(0, \infty)$. We have

$$f'(\ell) = \frac{1}{2}(1000 - 6\ell),$$ \hfill (4.67)

which is equal to zero when $\ell = 1000/6$. To confirm that this is the absolute maximum, we compute $f''(\ell)$ and obtain

$$f''(\ell) = \frac{1}{2}(-6) = -3.$$ \hfill (4.68)

Since $f''(\ell) < 0$, it follows that $f'(\ell)$ is always decreasing. Since $f'(1000/6) = 0$, it then follows that $f'(\ell) > 0$ for $0 < \ell < 1000/6$, and $f'(\ell) < 0$ for $\ell > 1000/6$. This, in turn, implies that $f(\ell)$ is increasing for $0 < \ell < 1000/6$, and decreasing for $\ell > 1000/6$. Therefore, by the First Derivative Test for Absolute Extrema, the absolute maximum of $f(\ell)$ occurs at $\ell = 1000/6$. 
We conclude that the dimensions of the rectangular region should be
\[ \ell = \frac{1000}{6} \text{ m}, \quad w = \frac{1000 - 3\ell}{2} = \frac{1000 - 3(1000/6)}{2} = 250 \text{ m}. \] (4.69)

Example 4.27 Find the shortest distance between the point \((-1, -1)\) and the line \(y = 4x + 2\).

**Solution** We proceed by following the general procedure for solving optimization problems.

1. **What is the objective function?** We wish to minimize the distance between the point \((-1, -1)\) and the line \(y = 4x + 2\). In other words, we wish to find the point \((x, y)\) on this line such that the distance between \((x, y)\) and \((-1, -1)\) is a minimum. Therefore, by the distance formula, our objective function is
   \[ f(x, y) = \sqrt{(x + 1)^2 + (y + 1)^2}. \] (4.70)

2. **What are the constraints on the variables?** Since the point \((x, y)\) lies on the line \(y = 4x + 2\), the equation \(y = 4x + 2\) defines a constraint on the values of \(x\) and \(y\).

3. **Eliminate variables until so that the objective function depends on only one variable.** We use the equation \(y = 4x + 2\) to eliminate \(y\) and obtain
   \[ f(x) = \sqrt{(x + 1)^2 + ((4x + 2) + 1)^2} = \sqrt{(x + 1)^2 + (4x + 3)^2} = \sqrt{17x^2 + 26x + 10}. \] (4.71)

4. **Find the minimum value of the objective function.** We wish to find the absolute minimum of \(f(x)\) on \((-\infty, \infty)\). As in the previous examples, we will use the First Derivative Test for Absolute Extrema, which requires us to determine where \(f\) is increasing and decreasing. To accomplish this, we find all critical numbers of \(f(x)\) on \((-\infty, \infty)\). We have
   \[ f'(x) = \frac{1}{2}(17x^2 + 26x + 10)^{-1/2}(34x + 26) = \frac{17x + 13}{\sqrt{17x^2 + 26x + 10}}. \] (4.72)

Therefore, \(f'(x) = 0\) only when \(x = -13/17\). By substituting other \(x\)-values into \(f'(x)\), we can determine that \(f'(x) > 0\) for \(x > -13/17\), and \(f'(x) < 0\) for \(x < -13/17\). This implies that \(f(x)\) is decreasing for \(x < -13/17\) and increasing for \(x > -13/17\).
4.5. OPTIMIZATION

−13/17, and increasing for $x > -13/17$. By the First Derivative Test for Absolute Extrema, we can conclude that the absolute minimum of $f(x)$ occurs at $x = -13/17$.

Using the constraint $y = 4x + 2$, we obtain

$$y = 4(-13/17) + 2 = -52/17 + 2 = -18/17.$$ \hfill (4.73)

We conclude that the shortest distance between the point $(-1, -1)$ and the line $y = 4x + 2$ is

$$\sqrt{(-13/17 + 1)^2 + (4(-18/17) + 2)^2} = \sqrt{(4/17)^2 + (-38/17)^2} = \frac{\sqrt{4^2 + 38^2}}{17} \approx 2.2476.$$ \hfill (4.74)

\[\square\]

**Example 4.28** Suppose that an isosceles triangle is inscribed in a circle of radius $r$. Prove that the area of the triangle is maximized when it is an equilateral triangle.

**Solution** The inscribed triangle is shown in Figure 4.16. For simplicity, we assume that the circle is centered at the origin $(0,0)$, and that one vertex of the triangle lies at the point $(0,r)$. Furthermore, we assume that this vertex is the one shared by the two sides of equal length. It follows that the remaining side, which is also the base, is a horizontal line segment. Therefore, the sides of the triangle, and therefore its area, are determined by the $y$-coordinate of the base.

We now follow the standard procedure for solving optimization problems. Our goal is to maximize the area of the triangle, given the constraints on its dimensions described in the problem statement, and show that the triangle of maximum area is in fact an equilateral triangle.

1. **What is the objective function?** We wish to maximize the area $A$ of the triangle, so our objective function is

$$A = f(b, h) = \frac{1}{2}bh,$$ \hfill (4.75)

where $b$ the base of the triangle and $h$ is the height.

2. **What are the constraints on the variables?** We can define both the base and the height in terms of the radius $r$ and the $y$-coordinate of the base, which we denote by $y$. The height is given by

$$h = r - y,$$ \hfill (4.76)
Figure 4.16: Isoceles triangle inscribed in a circle of radius $r$

since the height is the distance between the point $(0, r)$ and the point $(0, y)$. The base is given by

$$b = 2\sqrt{r^2 - y^2},$$

(4.77)

by the Pythagorean Theorem, as can be seen in Figure 4.16. We can see from the figure that $y$ must lie in the interval $[-r, r]$.

3. **Eliminate variables so that the objective function depends on only one variable.** In this case, it is actually easiest to eliminate both $h$ and $b$ and define the objective function as a function of $y$. From our definitions of $b$ and $h$ in terms of $y$, we obtain

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

(4.78)
4. **Find the maximum value of the objective function.** We now find the maximum value of \( f(y) \) on the interval \(-r \leq y \leq r\). We begin by finding any critical numbers of \( r \), which requires finding the points at which \( f'(y) = 0 \). We have, by the Product Rule,

\[
f'(y) = \frac{d}{dy} \left[ \sqrt{r^2 - y^2} (r - y) \right]
\]

\[
= (r - y) \frac{d}{dy} \left[ \sqrt{r^2 - y^2} \right] + \sqrt{r^2 - y^2} \frac{d}{dy} [(r - y)]
\]

\[
= (r - y) \frac{1}{2} (r^2 - y^2)^{-1/2} (-2y) + \sqrt{r^2 - y^2} (-1)
\]

\[
= y(y - r)(r^2 - y^2)^{-1/2} - \sqrt{r^2 - y^2}
\]

\[
= \sqrt{r^2 - y^2} \left[ \frac{y(y - r)}{r^2 - y^2} - 1 \right]
\]

\[
= \sqrt{r^2 - y^2} \left[ \frac{-y}{r + y} - 1 \right]
\]

\[
= \sqrt{r^2 - y^2} \left[ \frac{-y}{r + y} - \frac{r + y}{r + y} \right]
\]

\[
= \sqrt{r^2 - y^2} \left[ \frac{r + 2y}{r + y} \right].
\]

It follows that \( f'(y) = 0 \) when \( y = -r/2 \).

Since this is the only critical number of \( f(y) \) on the interval \([-r, r]\), it follows that the only points at which \( f \) can have an absolute minimum are at \( y = -r/2 \) and at the endpoints \( y = r \) and \( y = -r \). We therefore evaluate \( f(y) \) at these three points, and obtain

\[
f(-r/2) = \sqrt{r^2 - (-r/2)^2} (r - (-r/2))
\]

\[
= \sqrt{r^2 - r^2/4} (3r/2)
\]

\[
= \sqrt{3r^2/4} (3r/2)
\]

\[
= (\sqrt{3}r/2)(3r/2)
\]

\[
= \frac{3\sqrt{3}r^2}{4},
\]

\[
f(-r) = \sqrt{r^2 - (-r)^2} (r - (-r)) = \sqrt{r^2 - r^2}(2r) = 0,
\]

\[
(4.79)
\]
and
\[ f(r) = \sqrt{r^2 - r^2(r - r)} = 0. \] (4.80)
Therefore, the area is maximized when \( y = -r/2 \).

We still need to show that this value of \( y \) implies that the triangle is equilateral. We have
\[ b = 2\sqrt{r^2 - y^2} = 2\sqrt{r^2 - (-r/2)^2} = 2\sqrt{3r^2/4} = 2(\sqrt{3}r/2) = \sqrt{3}r, \] (4.81)
and
\[ h = r - y = \frac{3r}{2}. \] (4.82)
By the Pythagorean Theorem, the two equal sides of the triangle with common vertex at the point \((0, r)\) are each the hypotenuse of the right triangle with sides \( h \) and \( b/2 \). Let \( \ell \) denote the length of each of these sides. Then we have
\[
\ell = \sqrt{h^2 + (b/2)^2} \\
= \sqrt{\left(\frac{3r}{2}\right)^2 + \left(\frac{\sqrt{3}r}{2}\right)^2} \\
= \sqrt{\frac{9r^2}{4} + \frac{3r^2}{4}} \\
= \sqrt{\frac{12r^2}{4}} \\
= \sqrt{3}r.
\]
Since each of these sides has length equal to that of the base, the triangle must be equilateral. \( \Box \)

### 4.6 Newton’s Method

In finding the maximum or minimum value of a function on an interval, it is necessary to find the points at which the derivative is equal to zero. To this point, we have taken this part of the process for granted, but this can be a very difficult problem to solve. More often than not, it is actually impossible to solve the equation \( f(x) = 0 \) exactly for a given function \( f(x) \). In such cases, it is necessary to approximate the solution.
We have already seen one method, called \textit{bisection}, that can be used to solve the equation $f(x) = 0$. Given an interval $[a, b]$ on which $f$ is continuous, and changes sign, bisection is \textit{guaranteed} to find an approximate solution to $f(x) = 0$ that is as accurate as we wish. However, this approach has two drawbacks: first, it is rather slow, as several bisections of the interval $[a, b]$ may be necessary to obtain a sufficiently accurate solution. Second, it may not be simple to find an interval $[a, b]$ on which $f$ changes sign.

In this section, we will discuss an alternative approach, called \textit{Newton’s Method}, that overcomes these drawbacks. It finds an approximate solution to the equation $f(x) = 0$ without requiring knowledge of an interval $[a, b]$ on which a solution is guaranteed to exist, and it is usually much faster than bisection.

The basic idea behind Newton’s method in solving the equation $f(x) = 0$ is to approximate $f(x)$ by a linear function $g(x)$, since it is very easy to determine where a linear function crosses the $x$-axis; that is, where its $y$-value is equal to zero. To use this approach, two questions must be answered:

- What linear function $g(x)$ should be used to approximate $f(x)$?
- Given the approximate solution obtained by determining where $g(x) = 0$, how can this approximate solution be improved?

To answer the first question, we note that at any particular point $x_0$, $f(x)$ can be approximated fairly accurately by its tangent line at $x_0$, at least at points near $x_0$. Therefore, we can begin by choosing an arbitrary point $x_0$ to be our initial guess of the solution to $f(x) = 0$, obtain the equation of the tangent line to $f(x)$ at $x_0$, and then find the point $x_1$ at which this tangent line crosses the $x$-axis, which is the line $y = 0$. Since the tangent line has the equation

$$y - f(x_0) = f'(x_0)(x - x_0),$$

we can determine where this line crosses the $x$-axis by setting $y = 0$ in this equation and solving for $x$, which yields

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Let $x_1$ be this $x$-value at which the tangent line crosses the $x$-axis; that is, let

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$
CHAPTER 4. APPLICATIONS OF DERIVATIVES

If \( f(x_1) \) is very close to zero, then we could conclude that \( x_1 \) is a decent approximation to the solution of \( f(x) = 0 \) and stop. However, what if this is not the case? We can then repeat the process, computing the equation of the tangent line of \( f(x) \) at \( x_1 \) and determining the point \( x_2 \) at which this tangent line crosses the \( x \)-axis. This yields

\[
x_2 = x_1 - \frac{f(x_1)}{f'(x_1)},
\]

and therefore \( x_2 \) is yet another approximate solution to \( f(x) = 0 \).

We can continue this process as many times as we wish, until we decide that our latest approximate solution is a good enough approximation, in some sense. This iterative process is Newton’s Method. We select a starting iterate \( x_0 \), and compute \( x_1, x_2, x_3, \ldots \), using the iteration

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

until we have either obtained a sufficiently accurate approximate solution to \( f(x) = 0 \), or until it is apparent that the sequence of iterates \( x_1, x_2, \ldots \) is not converging to any particular value. In this situation, we need to choose a new starting iterate \( x_0 \), we we will discuss further below. First, however, we illustrate the use of Newton’s method with an example for which the sequence of iterates does converge to a solution.

**Example 4.29** We will use of Newton’s Method in computing \( \sqrt{2} \). This number satisfies the equation \( f(x) = 0 \) where

\[
f(x) = x^2 - 2.
\]

Since \( f'(x) = 2x \), it follows that in Newton’s Method, we can obtain the next iterate \( x_{n+1} \) from the previous iterate \( x_n \) by

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

We choose our starting iterate \( x_0 = 1 \), and compute the next several iterates as follows:

\[
x_1 = \frac{1}{2} + \frac{1}{1} = 1.5
\]

\[
x_2 = \frac{1.5}{2} + \frac{1}{1.5} = 1.4166667
\]

\[
x_3 = 1.41421569
\]

\[
x_4 = 1.41421356
\]

\[
x_5 = 1.41421356.
\]
Since the fourth and fifth iterates agree in to eight decimal places, we assume that 1.41421356 is a correct solution to $f(x) = 0$, to at least eight decimal places. The first two iterations are illustrated in Figure 4.17.

Figure 4.17: Newton’s Method applied to $f(x) = x^2 - 2$. The bold curve is the graph of $f$. The initial iterate $x_0$ is chosen to be 1. The tangent line of $f(x)$ at the point $(x_0, f(x_0))$ is used to approximate $f(x)$, and it crosses the $x$-axis at $x_1 = 1.5$, which is much closer to the exact solution than $x_0$. Then, the tangent line at $(x_1, f(x_1))$ is used to approximate $f(x)$, and it crosses the $x$-axis at $x_2 = 1.416$, which is already very close to the exact solution.

**Example 4.30** Newton’s Method can be used to compute the reciprocal of a number $a$ without performing any divisions. The solution, $1/a$, satisfies the equation $f(x) = 0$, where

$$f(x) = a - \frac{1}{x}.$$ (4.90)
Since
\[ f'(x) = \frac{1}{x^2}, \]  
(4.91)
it follows that in Newton’s Method, we can obtain the next iterate \( x_{n+1} \) from the previous iterate \( x_n \) by
\[ x_{n+1} = x_n - \frac{a - 1/x_n}{1/x_n^2} = x_n - \frac{a}{1/x_n} \frac{1/x_n}{1/x_n^2} = 2x_n - ax_n^2. \]  
(4.92)

Note that no divisions are necessary to obtain \( x_{n+1} \) from \( x_n \). This iteration was actually used on older IBM computers to implement division in hardware.

We use this iteration to compute the reciprocal of \( a = 8 \). Choosing our starting iterate to be 0.1, we compute the next several iterates as follows:

\[
\begin{align*}
  x_1 &= 2(0.1) - 8(0.1)^2 = 0.12 \\
  x_2 &= 2(0.12) - 8(0.12)^2 = 0.1248 \\
  x_3 &= 0.12499968 \\
  x_4 &= 0.12499999 \\
  x_5 &= 0.125 \\
  x_6 &= 0.125.
\end{align*}
\]

Since the fifth and sixth iterates agree, we assume that 0.125 is the correct solution.

Now, suppose we repeat this process, but with an initial iterate of \( x_0 = 1 \). Then, we have

\[
\begin{align*}
  x_1 &= 2(1) - 8(1)^2 = -6 \\
  x_2 &= 2(-6) - 8(-6)^2 = -300 \\
  x_3 &= 2(-300) - 8(-300)^2 = -720600
\end{align*}
\]

It is clear that this sequence of iterates is not going to converge to the correct solution. In general, for this iteration to converge to the reciprocal of \( a \), the initial iterate \( x_0 \) must be chosen so that \( 0 < x_0 < 2/a \). This condition guarantees that the next iterate \( x_1 \) will at least be positive. The contrast between the two choices of \( x_0 \) are illustrated in Figure 4.18. □

The preceding example shows that Newton’s Method has its drawbacks. In particular, it is important to choose the initial iterate \( x_0 \) properly, or the iteration may not converge to a solution at all. Unfortunately, there
4.6. NEWTON’S METHOD

Figure 4.18: Newton’s Method used to compute the reciprocal of 8 by solving the equation \( f(x) = 8 - 1/x = 0 \). When \( x_0 = 0.1 \), the tangent line of \( f(x) \) at \((x_0, f(x_0))\) crosses the \( x \)-axis at \( x_1 = 0.12 \), which is close to the exact solution. When \( x_0 = 1 \), the tangent line crosses the \( x \)-axis at \( x_1 = -6 \), which causes searching to continue on the wrong portion of the graph, so the sequence of iterates does not converge to the correct solution.

are no concrete guidelines for choosing \( x_0 \) so that Newton’s method will converge. By contrast, the bisection method, although it converges much more slowly, is guaranteed to converge if an interval containing a solution is known, no matter how large that interval is. Some methods for solving \( f(x) = 0 \) combine these two methods in some way, in order to take advantage of the desirable features of both.
4.7 Antiderivatives

Suppose that an object is moving in a straight line, and its acceleration is known; that is, we know that at time \( t \), its acceleration is \( a(t) \). This may be the case, for instance, if the only force acting on the object is gravity, so its acceleration, in \( \text{ft}/\text{s}^2 \), is given by \( a(t) = -32 \). How can we determine the object’s velocity as a function of time?

We can use the fact that acceleration is defined to be the rate of change of velocity with respect to time. In the notation of derivatives, this means that if \( v(t) \) is the velocity, then \( a(t) = v'(t) \). Therefore, we need to find a function whose derivative is equal to \( a(t) \). This involves reversing the process of differentiation. Previously, we would compute the derivative of a given function \( f(x) \); the function \( f \) was the input and its derivative \( f' \) was the output. In this situation, the derivative \( v' \) is the input and the original function \( v \) is the output. The process of recovering the original function from its derivative is called \textit{anti-differentiation}. The original function that is recovered from the given derivative is called an \textit{antiderivative}.

The process of anti-differentiation is complicated by the fact that the antiderivative of a function is not unique. To see this, suppose that we are able to find an antiderivative of our given acceleration function \( a(t) \); that is, we obtain a function \( v(t) \) such that \( v'(t) = a(t) \). Then, if we define \( u(t) = v(t) + C \), where \( C \) is any constant, then we have

\[
u'(t) = \frac{d}{dt}[v(t) + c] = v'(t) + \frac{d}{dt}[C] = a(t) + 0 = a(t).
\]

In other words, \( u(t) \) and \( v(t) \) are both antiderivatives of \( a(t) \). Since the constant \( C \) is arbitrary, we can conclude that \( a(t) \) has \textit{infinitely many} antiderivatives, since \( v(t) + C \) is an antiderivative for \textit{any} constant \( C \).

Fortunately, we can show that \textit{all} antiderivatives of \( a(t) \) are of the form \( v(t) + C \), where \( v(t) \) is any particular antiderivative of \( a(t) \) and \( C \) is a constant. This can be seen from a consequence of the Mean Value Theorem that was stated and proved in Section 4.2: any two functions that have the same derivative differ by a constant. From this result, we can conclude that if we are able to find \textit{one} function \( v(t) \) that is an antiderivative of \( a(t) \), then we can describe \textit{all} antiderivatives of \( a(t) \): they have the form \( v(t) + C \), where \( C \) is a constant.

It follows from this discussion that the task of anti-differentiating a given function \( f(x) \) consists of two main tasks:

- How can we find a single function \( F(x) \) such that \( F'(x) = f(x) \)?
4.7. **ANTIDERIVATIVES**

- Since any function of the form $F(x) + C$ is also an antiderivative of $f(x)$, how can we determine the correct value of $C$?

The first task can be accomplished by reversing known differentiation rules, as the following examples illustrate.

**Example 4.31** Let $f(x) = \cos x$. We know that

$$\frac{d}{dx}[\sin x] = \cos x,$$  \hspace{1cm} (4.94)

so $\sin x$ is an antiderivative of $\cos x$. Similarly, because

$$\frac{d}{dx}[\cos x] = -\sin x,$$  \hspace{1cm} (4.95)

we can conclude that an antiderivative of $\sin x$ is $-\cos x$. To confirm this, note that

$$\frac{d}{dx}[-\cos x] = -\frac{d}{dx}[\cos x] = -(-\sin x) = \sin x.$$  \hspace{1cm} (4.96)

\[\square\]

**Example 4.32** Let $f(x) = 1$. From the basic rule

$$\frac{d}{dx}[x] = 1,$$  \hspace{1cm} (4.97)

we can see that $x$ is an antiderivative of $1$. Similarly, if we start with the relation

$$\frac{d}{dx}[x^2] = 2x,$$  \hspace{1cm} (4.98)

and divide both sides by 2, we have

$$\frac{d}{dx}\left[\frac{x^2}{2}\right] = x,$$  \hspace{1cm} (4.99)

so $x^2/2$ is an antiderivative of $x$. More generally, we can divide the Power Rule

$$\frac{d}{dx}[x^n] = nx^{n-1}$$  \hspace{1cm} (4.100)

by $n$, assuming $n \neq 0$, and obtain

$$\frac{d}{dx}\left[\frac{x^n}{n}\right] = x^{n-1},$$ \hspace{1cm} (4.101)

which implies that $x^n/n$ is an antiderivative of $x^{n-1}$, if $n \neq 0$. \[\square\]
From many of the differentiation rules we have learned, we can obtain rules for computing antiderivatives. These rules are called *anti-differentiation rules*, but, for reasons that will become clear upon learning integral calculus, they are more commonly known as *integration rules*. In the following table, we use the notation that $F$ is an antiderivative of $f$ and $G$ is an antiderivative of $g$; that is, $F' = f$ and $G' = g$.

<table>
<thead>
<tr>
<th>Function</th>
<th>Antiderivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$cf(x)$</td>
<td>$cF(x)$</td>
</tr>
<tr>
<td>$f(x) + g(x)$</td>
<td>$F(x) + G(x)$</td>
</tr>
<tr>
<td>$x^n$</td>
<td>$\frac{x^{n+1}}{n+1}$, $n \neq -1$</td>
</tr>
<tr>
<td>$\cos x$</td>
<td>$\sin x$</td>
</tr>
<tr>
<td>$\sin x$</td>
<td>$-\cos x$</td>
</tr>
<tr>
<td>$\sec^2 x$</td>
<td>$\tan x$</td>
</tr>
<tr>
<td>$\sec x \tan x$</td>
<td>$\sec x$</td>
</tr>
</tbody>
</table>

Using these rules, we can obtain a particular antiderivative of a given function $f(x)$, in at least some cases. Many more anti-differentiation rules can be found in any text on integral calculus.

Once we have obtained a particular antiderivative $F(x)$ of a given function $f(x)$, how can we determine the *specific* antiderivative for the problem at hand? We know that any antiderivative of $f(x)$ is of the form $F(x) + C$, where $C$ is a constant, so we need to determine the value of $C$ that is appropriate for the given problem. For example, suppose that we have anti-differentiated our acceleration $a(t)$ from the preceding discussion to obtain a function $v(t)$. Let $V(t)$ be the *actual* velocity at time $t$ of the object whose acceleration is $a(t)$. Then, we know that $V(t) = v(t) + C$, for some constant $C$. To determine the value of $C$, we can measure the velocity at a *specific* time $t_0$. Then, we have $V(t_0) = v(t_0) + C$, and since $V(t_0)$ and $v(t_0)$ are both known, it follows that the correct value of $C$ for the given problem is $V(t_0) - v(t_0)$.

In general, if we are given a function $f(x)$ that is the derivative of an unknown function $G(x)$, and we need to determine $G(x)$, we first need to compute a particular antiderivative $F(x)$, and then we know that $G(x) = F(x) + C$ for some constant $C$. If we know the value of $G(x)$ at some point $x_0$, then the value of $C$ is given by $C = G(x_0) - F(x_0)$. In many applications, the point $x_0$ at which the value of $G(x)$ is known corresponds to some “initial” point, such as an initial time, and in such cases the value $G(x_0)$ is called an *initial value*. The problem of obtaining a function from its derivative and an initial value is called an *initial value problem*, which
is one of the most basic types of differential equations. We illustrate the solution of simple initial value problems in the following example.

**Example 4.33** A ball is thrown into the air from a height of 40 ft, with an initial velocity of 60 ft/s. What is the function \( s(t) \) that describes the height of the ball at time \( t \)?

**Solution** The acceleration of the ball at time \( t \) is described by the very simple function \( a(t) = -32 \text{ ft/s}^2 \), since the acceleration is affected only by gravity. Since acceleration is the second derivative of position, we must anti-differentiate \( a(t) \) twice in order to obtain \( s(t) \).

First, we anti-differentiate \( a(t) \) to obtain the velocity, which we denote by \( v(t) \). Using the rule that an antiderivative of \( x^n \) is \( x^{n+1}/(n+1) \), and that an antiderivative of \( cf(x) \) is \( c \) times the antiderivative of \( f(x) \), we find that an antiderivative of \(-32\) is \(-32t\). Therefore, \( v(t) = -32t + C \), for some constant \( C \) that needs to be determined. Since the initial velocity of the ball is 60 ft/s, this implies that \( v(0) = 60 \). Substituting \( t = 0 \) into the relation \( v(t) = -32t + C \), we conclude that the correct value of \( C \) is 60, and therefore \( v(t) = -32t + 60 \).

Now, we anti-differentiate the velocity to obtain the position \( s(t) \). Using the same anti-differentiation rules as before, we obtain \( s(t) = -32t^2/2 + 60t + C \), or \( s(t) = -16t^2 + 60t + C \), for some constant \( C \). The fact that the initial height of the ball is 40 ft implies that \( s(0) = 40 \). Substituting \( t = 0 \) into the relation \( s(t) = -16t^2 + 60t + C \), we can conclude that \( C = 40 \), so the position at time \( t \) is given by \( s(t) = -16t^2 + 60t + 40 \).

The anti-differentiation rules we have seen so far are only useful for a limited selection of functions. When it is not possible to use these rules to obtain the antiderivative, it is still possible to obtain an approximate graph of the antiderivative. Suppose we wish to compute the antiderivative \( F(x) \) of a given function \( f(x) \), and we know that \( F(x_0) = y_0 \). We begin by plotting the point \((x_0, y_0)\). Then, we use the fact that \( F'(x) = f(x) \) to obtain the slope of the tangent line of \( F(x) \) at the point \((x_0, y_0)\). This tangent line can provide the approximate value of \( F(x) \) at a point \( x_1 \) that is near \( x_0 \). Then, we use the tangent line of \( F(x) \) at \( x_1 \), which has slope \( f(x_1) \), to obtain the value of \( F(x) \) at yet another point \( x_2 \), and so on. Continuing in this fashion, we can obtain a graph of a function that is approximately equal to \( F(x) \).
Index

anti-differentiation, 248
anti-differentiation rules, 250
antiderivative, 248
astroid, 125
asymptote, horizontal, 66
asymptote, vertical, 62, 65
basis step, 93, 210
bisection, 59

cancellation equations, 162, 185
Chain Rule, 105
closed interval method, 201
coefficient, of polynomial, 14
composition, 17, 57
compressibility, 102
concave downward, 214
concave upward, 214
concavity, 214
Concavity Test, 214
conjugate, 48, 70
Constant Multiple Rule, 87
continuity, 51
continuity from left, 55
continuity from right, 55
Continuity Laws, 56
continuity on interval, 56
continuity, one-sided, 55
converse, 84, 199
critical number, 198
degree, of polynomial, 14
density, 102
dependent variable, 13
derivative, 79
derivative, $n$th, 115
derivative, as function, 80
derivative, second, 113
derivative, third, 114
difference quotient, 77
Difference Rule, 87
differentiable, 79
differential calculus, 9
differential equation, 180, 251
differentiation, 83
differentiation operator, 83
differentiation rules, 85
differentiation, implicit, 122
Direct Substitution Property, 42, 43
discontinuity, 52
discontinuity, infinite, 54
discontinuity, jump, 54
discontinuity, removable, 54
domain, 157
equation, explicit, 122
equation, implicit, 122
extreme value, 197
Extreme Value Theorem, 198
extremum, 197
extremum, local, 198
Fermat’s Theorem, 198
First Derivative Test, 212
First Derivative Test for Absolute Ex-
treme Value, 234
function, algebraic, 14
function, constant, 39, 85
function, cubic, 14
function, even, 89, 221
function, identity, 39, 86
function, inverse, 157
function, linear, 13
function, polynomial, 14
function, power, 14
function, quadratic, 14
function, rational, 14
function, square root, 14
function, trigonometric, 15
greatest integer function, 51
Heaviside function, 37
horizontal line test, 159
Increasing/Decreasing Test, 211
increment, 77
independent variable, 13
indeterminate form, 190, 210
induction step, 94, 210
inflection point, 214
initial condition, 180
initial value, 180, 250
initial value problem, 180, 250
Intermediate Value Theorem, 59
inverse sine function, 184
jerk, 113
l'Hospital's Rule, 190
law of natural decay, 181
law of natural growth, 181
limit, 11, 31, 33
limit at infinity, 65
limit laws, 40
limit, infinite, 61, 68
limit, left-hand, 39
limit, right-hand, 39
logarithmic differentiation, 176
marginal cost, 102
marginal demand, 103
marginal profit, 103
marginal revenue, 103
mathematical induction, 93, 210
maximum value, 197
maximum, absolute, 197
maximum, global, 197
maximum, local, 198
Mean Value Theorem, 205
minimum value, 197
minimum, absolute, 197
minimum, global, 197
minimum, local, 198
model, 12
Newton's Method, 243
objective function, 197
one-to-one, 159
optimization, 197
optimization problem, 228
order of derivative, 115
period, 219
periodicity, 219
point-slope form, 24
Power Rule, 92
Power Rule, General, 97
problem, forward, 157
problem, inverse, 157
Product Rule, 91
Quotient Rule, 96
range, 157
rate of change, average, 9, 77
rate of change, instantaneous, 9, 77
rate of reaction, 102
related rates, 127
relative growth rate, 181
Rolle’s Theorem, 204

secant line, 18
Second Derivative Test, 215
slope, 13
slope-intercept form, 13, 25
Squeeze Theorem, 43
Sum Rule, 86

tangency, point of, 19
tangent, 19
tangent line, 19, 75

unit circle, 122

velocity, average, 9
velocity, instantaneous, 10, 73

y-intercept, 13