MAT 2420 CALCULUS II



LibreTexts™

MAT 202 Calculus II

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide and accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 06/08/2025

TABLE OF CONTENTS

Licensing

2: Applications of Integration

- 2.E: Applications of Integration (Exercises)
- 2.0: Prelude to Applications of Integration
- 2.1: Areas between Curves
- 2.2: Determining Volumes by Slicing
- 2.3: Volumes of Revolution - Cylindrical Shells
- 2.4: Arc Length of a Curve and Surface Area
- 2.5: Physical Applications of Integration
- 2.6: Moments and Centers of Mass
- 2.7: Integrals, Exponential Functions, and Logarithms
- 2.8: Exponential Growth and Decay
- 2.9: Calculus of the Hyperbolic Functions

3: Techniques of Integration

- 3.E: Techniques of Integration (Exercises)
- 3.0: Prelude to Techniques of Integration
- 3.1: Integration by Parts
- 3.2: Trigonometric Integrals
- 3.3: Trigonometric Substitution
- 3.4: Partial Fractions
- 3.5: Other Strategies for Integration
- 3.6: Numerical Integration
- 3.7: Improper Integrals

4: Introduction to Differential Equations

- 4.E: Introduction to Differential Equations (Exercises)
- 4.0: Prelude to Differential Equations
- 4.1: Basics of Differential Equations
- 4.2: Direction Fields and Numerical Methods
- 4.3: Separable Equations
- 4.4: The Logistic Equation

5: Sequences and Series

- 5.E: Sequences and Series (Exercises)
- 5.0: Prelude to Sequence and Series
- 5.1: Sequences
- 5.2: Infinite Series
- 5.3: The Divergence and Integral Tests
- 5.4: Comparison Tests
- 5.5: Alternating Series
- 5.6: Ratio and Root Tests

6: Power Series

- 6.E: Power Series (Exercises)
- 6.0: Prelude to Power Series
- 6.1: Power Series and Functions
- 6.2: Properties of Power Series
- 6.3: Taylor and Maclaurin Series
- 6.4: Working with Taylor Series

7: Parametric Equations and Polar Coordinates

- 7.E: Parametric Equations and Polar Coordinates (Exercises)
- 7.0: Prelude to Parametric Equations and Polar Coordinates
- 7.1: Parametric Equations
- 7.2: Calculus of Parametric Curves
- 7.3: Polar Coordinates
- 7.4: Area and Arc Length in Polar Coordinates
- 7.5: Conic Sections

Index

Glossary

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

2: Applications of Integration

In this chapter, we use definite integrals to calculate the force exerted on the dam when the reservoir is full and we examine how changing water levels affect that force. Hydrostatic force is only one of the many applications of definite integrals we explore in this chapter. From geometric applications such as surface area and volume, to physical applications such as mass and work, to growth and decay models, definite integrals are a powerful tool to help us understand and model the world around us.

2.E: Applications of Integration (Exercises)
2.0: Prelude to Applications of Integration
2.1: Areas between Curves
2.2: Determining Volumes by Slicing
2.3: Volumes of Revolution - Cylindrical Shells
2.4: Arc Length of a Curve and Surface Area
2.5: Physical Applications of Integration
2.6: Moments and Centers of Mass
2.7: Integrals, Exponential Functions, and Logarithms
2.8: Exponential Growth and Decay
2.9: Calculus of the Hyperbolic Functions

Thumbnail: A region between two functions.

Contributors

- Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 2: Applications of Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.E: Applications of Integration (Exercises)

This page titled 2.E: Applications of Integration (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.0: Prelude to Applications of Integration

The Hoover Dam is an engineering marvel. When Lake Mead, the reservoir behind the dam, is full, the dam withstands a great deal of force. However, water levels in the lake vary considerably as a result of droughts and varying water demands. Later in this chapter, we use definite integrals to calculate the force exerted on the dam when the reservoir is full and we examine how changing water levels affect that force. Hydrostatic force is only one of the many applications of definite integrals we explore in this chapter. From geometric applications such as surface area and volume, to physical applications such as mass and work, to growth and decay models, definite integrals are a powerful tool to help us understand and model the world around us.



Figure 2.0.1: Hoover Dam is one of the United States' iconic landmarks, and provides irrigation and hydroelectric power for millions of people in the southwest United States. (credit: modification of work by Lynn Betts, Wikimedia).

This page titled 2.0: Prelude to Applications of Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.0: Prelude to Applications of Integration** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.1: Areas between Curves

Learning Objectives

- Determine the area of a region between two curves by integrating with respect to the independent variable.
- Find the area of a compound region.
- Determine the area of a region between two curves by integrating with respect to the dependent variable.

In Introduction to Integration, we developed the concept of the definite integral to calculate the area below a curve on a given interval. In this section, we expand that idea to calculate the area of more complex regions. We start by finding the area between two curves that are functions of x, beginning with the simple case in which one function value is always greater than the other. We then look at cases when the graphs of the functions cross. Last, we consider how to calculate the area between two curves that are functions of y.

Area of a Region between Two Curves

Let $f(x)$ and $g(x)$ be continuous functions over an interval $[a, b]$ such that $f(x) \geq g(x)$ on $[a, b]$. We want to find the area between the graphs of the functions, as shown in Figure 2.1.1.

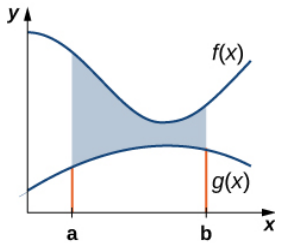


Figure 2.1.1: The area between the graphs of two functions, $f(x)$ and $g(x)$, on the interval $[a, b]$

As we did before, we are going to partition the interval on the **x-axis** and approximate the area between the graphs of the functions with rectangles. So, for $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of $[a, b]$. Then, for $i = 1, 2, \ldots, n$, choose a point $x_i^* \in [x_{i-1}, x_i]$, and on each interval $[x_{i-1}, x_i]$ construct a rectangle that extends vertically from $g(x_i^*)$ to $f(x_i^*)$. Figure 2.1.2a shows the rectangles when x_i^* is selected to be the left endpoint of the interval and $n = 10$. Figure 2.1.2b shows a representative rectangle in detail.

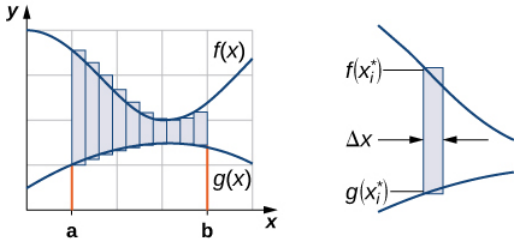


Figure 2.1.2: (a)We can approximate the area between the graphs of two functions, $f(x)$ and $g(x)$, with rectangles. (b) The area of a typical rectangle goes from one curve to the other.

The height of each individual rectangle is $f(x_i^*) - g(x_i^*)$ and the width of each rectangle is Δx. Adding the areas of all the rectangles, we see that the area between the curves is approximated by

$$A \approx \sum_{i=1}^{n} [f(x_i^*) - g(x_i^*)] \Delta x. \tag{2.1.1}$$

This is a Riemann sum, so we take the limit as $n \to \infty$ and we get

$$A = \lim_{n \to \infty} \sum_{i=1}^{n} [f(x_i^*) - g(x_i^*)] \Delta x = \int_a^b [f(x) - g(x)] dx. \tag{2.1.2}$$

These findings are summarized in the following theorem.

Finding the Area between Two Curves

Let $f(x)$ and $g(x)$ be continuous functions such that $f(x) \geq g(x)$ over an interval $[a, b]$. Let R denote the region bounded above by the graph of $f(x)$, below by the graph of $g(x)$, and on the left and right by the lines $x = a$ and $x = b$, respectively. Then, the area of $\mathbf{R}$ is given by

$$A = \int_a^b [f(x) - g(x)]dx. \tag{2.1.3}$$

We apply this theorem in the following example.

Example 2.1.1: Finding the Area of a Region between Two Curves I

If R is the region bounded above by the graph of the function $f(x) = x + 4$ and below by the graph of the function $g(x) = 3 - \frac{x}{2}$ over the interval $[1, 4]$, find the area of region $\mathbf{R}$.

Solution

The region is depicted in the following figure.

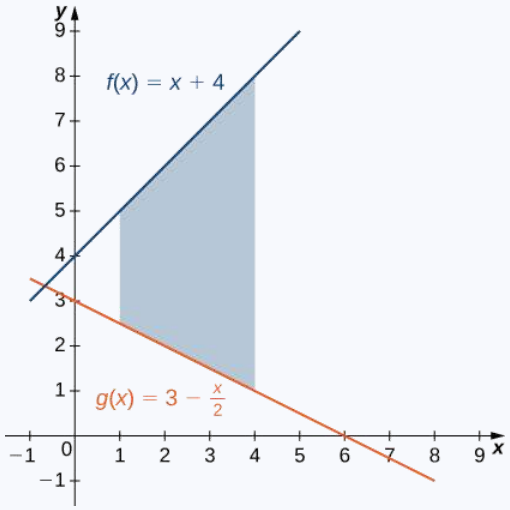


Figure 2.1.3: A region between two curves is shown where one curve is always greater than the other.

We have

$$A = \int_a^b [f(x) - g(x)]\, dx$$

$$= \int_1^4 [(x+4) - (3 - \frac{x}{2})]\, dx = \int_1^4 \left[\frac{3x}{2} + 1\right]\, dx$$

$$= [\frac{3x^2}{4} + x]\Big|_1^4 = (16 - \frac{7}{4}) = \frac{57}{4}.$$

The area of the region is $\frac{57}{4} units^2$.

Exercise 2.1.1

If $\mathbf{R}$ is the region bounded by the graphs of the functions $f(x) = \frac{x}{2} + 5$ and $g(x) = x + \frac{1}{2}$ over the interval $[1, 5]$, find the area of region $\mathbf{R}$.

Hint

Graph the functions to determine which function's graph forms the upper bound and which forms the lower bound, then follow the process used in Example.

Answer

$12\,\text{units}^2$

In Example, we defined the interval of interest as part of the problem statement. Quite often, though, we want to define our interval of interest based on where the graphs of the two functions intersect. This is illustrated in the following example.

Example 2.1.2: Finding the Area of a Region between Two Curves II

If $\mathbf{R}$ is the region bounded above by the graph of the function $f(x) = 9 - (x/2)^2$ and below by the graph of the function $g(x) = 6 - x$, find the area of region $\mathbf{R}$.

Solution

The region is depicted in the following figure.

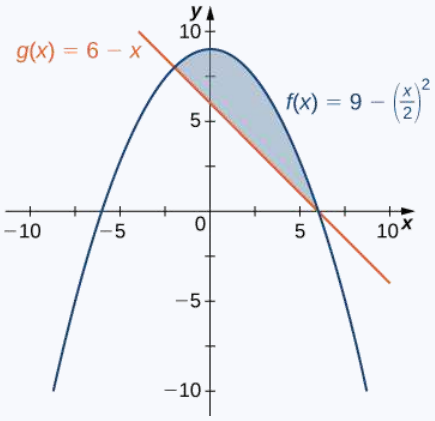


Figure 2.1.4: This graph shows the region below the graph of $f(x)$ and above the graph of $g(x)$.

We first need to compute where the graphs of the functions intersect. Setting $f(x) = g(x)$, we get

$$\begin{aligned} f(x) &= g(x) \\ 9 - (\frac{x}{2})^2 &= 6 - x \\ 9 - \frac{x^2}{4} &= 6 - x \\ 36 - x^2 &= 24 - 4x \\ x^2 - 4x - 12 &= 0 \\ (x-6)(x+2) &= 0. \end{aligned}$$

The graphs of the functions intersect when $x = 6$ or $x = -2$, so we want to integrate from -2 to 6. Since $f(x) \geq g(x)$ for $-2 \leq x \leq 6$, we obtain

$$\begin{aligned} A &= \int_a^b [f(x) - g(x)]\,dx \\ &= \int_{-2}^{6} \left[9 - (\frac{x}{2})^2 - (6 - x)\right]\,dx \\ &= \int_{-2}^{6} \left[3 - \frac{x^2}{4} + x\right]\,dx \\ &= \left[3x - \frac{x^3}{12} + \frac{x^2}{2}\right]\Big|_{-2}^{6} = \frac{64}{3}. \end{aligned}$$

The area of the region is $64/3\,\text{units}^2$.

Exercise 2.1.2

If R is the region bounded above by the graph of the function $f(x) = x$ and below by the graph of the function $g(x) = x^4$, find the area of region $\mathbf{R}$.

Hint

Use the process from Example 2.1.2

Answer

$\frac{3}{10}$ unit^2

Areas of Compound Regions

So far, we have required $f(x) \geq g(x)$ over the entire interval of interest, but what if we want to look at regions bounded by the graphs of functions that cross one another? In that case, we modify the process we just developed by using the absolute value function.

Finding the Area of a Region between Curves That Cross

Let $f(x)$ and $g(x)$ be continuous functions over an interval $[a, b]$. Let $\mathbf{R}$ denote the region between the graphs of $f(x)$ and $g(x)$, and be bounded on the left and right by the lines $x = a$ and $x = b$, respectively. Then, the area of $\mathbf{R}$ is given by

$$A = \int_a^b |f(x) - g(x)| dx. \tag{2.1.4}$$

In practice, applying this theorem requires us to break up the interval $[a, b]$ and evaluate several integrals, depending on which of the function values is greater over a given part of the interval. We study this process in the following example.

Example 2.1.3: Finding the Area of a Region Bounded by Functions That Cross

If R is the region between the graphs of the functions $f(x) = \sin x$ and $g(x) = \cos x$ over the interval $[0, \pi]$, find the area of region $\mathbf{R}$.

Solution

The region is depicted in the following figure.

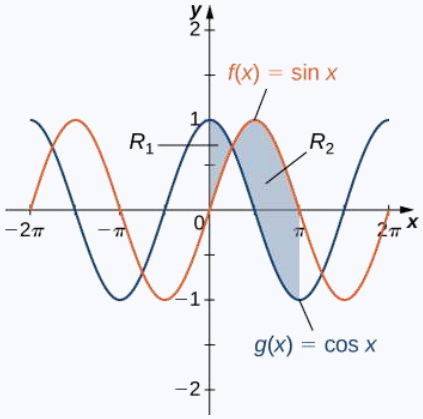


Figure 2.1.5: The region between two curves can be broken into two sub-regions.

The graphs of the functions intersect at $x = \pi/4$. For $x \in [0, \pi/4]$, $\cos x \geq \sin x$, so

$$|f(x) - g(x)| = |\sin x - \cos x| = \cos x - \sin x.$$

On the other hand, for $x \in [\pi/4, \pi]$, $\sin x \geq \cos x$,so

$$|f(x) - g(x)| = |\sin x - \cos x| = \sin x - \cos x.$$

Then

$$A = \int_a^b |f(x) - g(x)|dx$$

$$= \int_0^{\pi} |\sin x - \cos x|dx = \int_0^{\pi/4} (\cos x - \sin x)dx + \int_{\pi/4}^{\pi} (\sin x - \cos x)dx$$

$$= [\sin x + \cos x]|_0^{\pi/4} + [-\cos x - \sin x]|_{\pi/4}^{\pi}$$

$$= (\sqrt{2} - 1) + (1 + \sqrt{2}) = 2\sqrt{2}.$$

The area of the region is $2\sqrt{2}$ units^2.

Exercise 2.1.3

If R is the region between the graphs of the functions $f(x) = \sin x$ and $g(x) = \cos x$ over the interval $[\pi/2, 2\pi]$, find the area of region $\mathbf{R}$.

Hint

The two curves intersect at $x = (5\pi)/4$.

Answer

$2 + 2\sqrt{2}$ units^2

Example 2.1.4: Finding the Area of a Complex Region

Consider the region depicted in Figure 2.1.6. Find the area of $\mathbf{R}$.

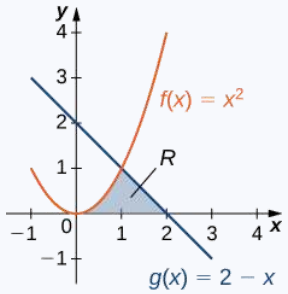


Figure 2.1.6: Two integrals are required to calculate the area of this region.

Solution

As with Example \PageIndex{3}, we need to divide the interval into two pieces. The graphs of the functions intersect at $x = 1$ (set $f(x) = g(x)$ and solve for x), so we evaluate two separate integrals: one over the interval $[0, 1]$ and one over the interval $[1, 2]$.

Over the interval $[0, 1]$, the region is bounded above by $f(x) = x^2$ and below by the x-axis, so we have

$$A_1 = \int_0^1 x^2 dx = \frac{x^3}{3}\,|_0^1 = \frac{1}{3}.$$

Over the interval $[1, 2]$, the region is bounded above by $g(x) = 2 - x$ and below by the **x-axis**, so we have

$$A_2 = \int_1^2 (2 - x)dx = [2x - \frac{x^2}{2}]\,|_1^2 = \frac{1}{2}.$$

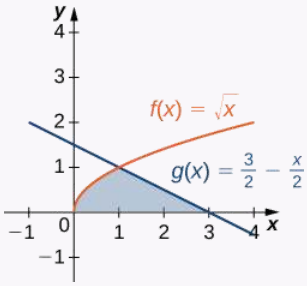
Adding these areas together, we obtain

$$A = A_1 + A_2 = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

The area of the region is $5/6$ units^2.

Exercise 2.1.4

Consider the region depicted in the following figure. Find the area of **R**.



Hint

The two curves intersect at **x=1**

Answer

$\frac{5}{3}$ units^2

Regions Defined with Respect to y

In Example, we had to evaluate two separate integrals to calculate the area of the region. However, there is another approach that requires only one integral. What if we treat the curves as functions of y, instead of as functions of x? Review Figure. Note that the left graph, shown in red, is represented by the function $y = f(x) = x^2$. We could just as easily solve this for x and represent the curve by the function $x = v(y) = \sqrt{y}$. (Note that $x = -\sqrt{y}$ is also a valid representation of the function $y = f(x) = x^2$ as a function of y. However, based on the graph, it is clear we are interested in the positive square root.) Similarly, the right graph is represented by the function $y = g(x) = 2 - x$, but could just as easily be represented by the function $x = u(y) = 2 - y$. When the graphs are represented as functions of y, we see the region is bounded on the left by the graph of one function and on the right by the graph of the other function. Therefore, if we integrate with respect to y, we need to evaluate one integral only. Let's develop a formula for this type of integration.

Let $u(y)$ and $v(y)$ be continuous functions over an interval $[c, d]$ such that $u(y) \geq v(y)$ for all $y \in [c, d]$. We want to find the area between the graphs of the functions, as shown in Figure 2.1.7.

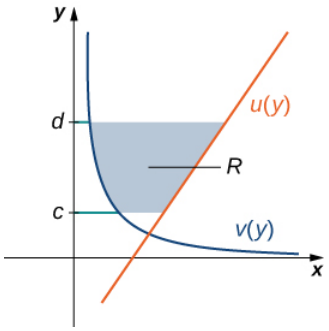
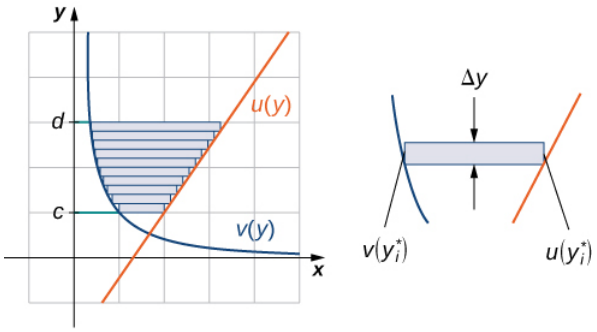


Figure 2.1.7: We can find the area between the graphs of two functions, $u(y)$ and $v(y)$.

This time, we are going to partition the interval on the **y-axis** and use horizontal rectangles to approximate the area between the functions. So, for $i = 0, 1, 2, \ldots, n$, let $Q = y_i$ be a regular partition of $[c, d]$. Then, for $i = 1, 2, \ldots, n$, choose a point $y_i^* \in [y_{i-1}, y_i]$, then over each interval $[y_{i-1}, y_i]$ construct a rectangle that extends horizontally from $v(y^0 *_i)$ to $u(y_i^*)$. Figure 2.1.8*a* shows the rectangles when y_i^* is selected to be the lower endpoint of the interval and $n = 10$. Figure 2.1.8*b* shows a representative rectangle in detail.



(a) (b)

Figure 2.1.8: (a) Approximating the area between the graphs of two functions, $u(y)$ and $v(y)$, with rectangles. (b) The area of a typical rectangle.

The height of each individual rectangle is Δy and the width of each rectangle is $u(y_i^*) - v(y_i^*)$. Therefore, the area between the curves is approximately

$$A \approx \sum_{i=1}^{n} [u(y_i^*) - v(y_i^*)] \Delta y.$$

This is a Riemann sum, so we take the limit as $n \to \infty$, obtaining

$$\begin{aligned} A &= \lim_{n \to \infty} \sum_{i=1}^{n} [u(y_i^*) - v(y_i^*)] \Delta y \\ &= \int_c^d [u(y) - v(y)] dy. \end{aligned}$$

These findings are summarized in the following theorem.

Finding the Area between Two Curves, Integrating along the y-axis

Let $u(y)$ and $v(y)$ be continuous functions such that $u(y) \geq v(y)$ for all $y \in [c, d]$. Let \(\textbf{R}\) denote the region bounded on the right by the graph of $u(y)$, on the left by the graph of $v(y)$, and above and below by the lines $y = d$ and $y = c$, respectively. Then, the area of $\mathbf{R}$ is given by

$$A = \int_c^d [u(y) - v(y)] dy. \tag{2.1.5}$$

Example 2.1.5: Integrating with Respect to y

Let's revisit Example, only this time let's integrate with respect to y. Let $\mathbf{R}$ be the region depicted in Figure 2.1.9. Find the area of $\mathbf{R}$ by integrating with respect to y.

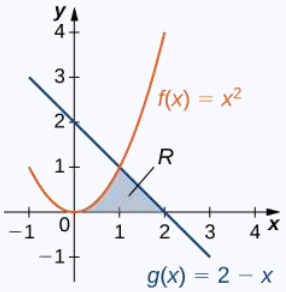


Figure 2.1.9: The area of region $\mathbf{R}$ can be calculated using one integral only when the curves are treated as functions of y.

Solution

We must first express the graphs as functions of y. As we saw at the beginning of this section, the curve on the left can be represented by the function $x = v(y) = \sqrt{y}$, and the curve on the right can be represented by the function $x = u(y) = 2 - y$.

Now we have to determine the limits of integration. The region is bounded below by the x-axis, so the lower limit of integration is $y = 0$. The upper limit of integration is determined by the point where the two graphs intersect, which is the point $(1, 1)$, so the upper limit of integration is $y = 1$. Thus, we have $[c, d] = [0, 1]$.

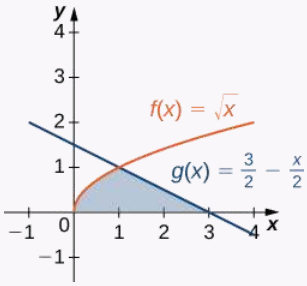
Calculating the area of the region, we get

$$\begin{aligned} A &= \int_c^d [u(y) - v(y)]dy \\ &= \int_0^1 [(2 - y) - \sqrt{y}]dy \\ &= [2y - \frac{y^2}{2} - \frac{2}{3}y^{3/2}]\Big|_0^1 \\ &= \frac{5}{6}. \end{aligned}$$

The area of the region is $5/6$ units^2.

Exercise 2.1.5

Let's revisit the checkpoint associated with Example, only this time, let's integrate with respect to y. Let be the region depicted in the following figure. Find the area of $\mathbf{R}$ by integrating with respect to y.



Hint

Follow the process from the previous example.

Answer

$\frac{5}{3}$ units^2

Key Concepts

- Just as definite integrals can be used to find the area under a curve, they can also be used to find the area between two curves.
- To find the area between two curves defined by functions, integrate the difference of the functions.
- If the graphs of the functions cross, or if the region is complex, use the absolute value of the difference of the functions. In this case, it may be necessary to evaluate two or more integrals and add the results to find the area of the region.
- Sometimes it can be easier to integrate with respect to y to find the area. The principles are the same regardless of which variable is used as the variable of integration.

Key Equations

- **Area between two curves, integrating on the x-axis**

  $$A = \int_a^b [f(x) - g(x)]dx$$

- **Area between two curves, integrating on the y-axis**

$$A = \int_c^d [u(y) - v(y)]dy$$

Contributors

- Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 2.1: Areas between Curves is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Determining Volumes by Slicing

Learning Objectives

- Determine the volume of a solid by integrating a cross-section (the slicing method).
- Find the volume of a solid of revolution using the disk method.
- Find the volume of a solid of revolution with a cavity using the washer method.

In the preceding section, we used definite integrals to find the area between two curves. In this section, we use definite integrals to find volumes of three-dimensional solids. We consider three approaches—slicing, disks, and washers—for finding these volumes, depending on the characteristics of the solid.

Volume and the Slicing Method

Just as area is the numerical measure of a two-dimensional region, volume is the numerical measure of a three-dimensional solid. Most of us have computed volumes of solids by using basic geometric formulas. The volume of a rectangular solid, for example, can be computed by multiplying length, width, and height: $V = lwh$. The formulas for the volumes of:

- a sphere

$$V_{sphere} = \frac{4}{3}\pi r^3, \tag{2.2.1}$$

- a cone

$$V_{cone} = \frac{1}{3}\pi r^2 h \tag{2.2.2}$$

- and a pyramid

$$V_{pyramid} = \frac{1}{3}Ah \tag{2.2.3}$$

have also been introduced. Although some of these formulas were derived using geometry alone, all these formulas can be obtained by using integration.

We can also calculate the volume of a cylinder. Although most of us think of a cylinder as having a circular base, such as a soup can or a metal rod, in mathematics the word cylinder has a more general meaning. To discuss cylinders in this more general context, we first need to define some vocabulary.

We define the *cross-section* of a solid to be the intersection of a plane with the solid. A cylinder is defined as any solid that can be generated by translating a plane region along a line perpendicular to the region, called the axis of the cylinder. Thus, all cross-sections perpendicular to the axis of a cylinder are identical. The solid shown in Figure 2.2.1 is an example of a cylinder with a noncircular base. To calculate the volume of a cylinder, then, we simply multiply the area of the cross-section by the height of the cylinder: $V = A \cdot h$. In the case of a right circular cylinder (soup can), this becomes $V = \pi r^2 h$.

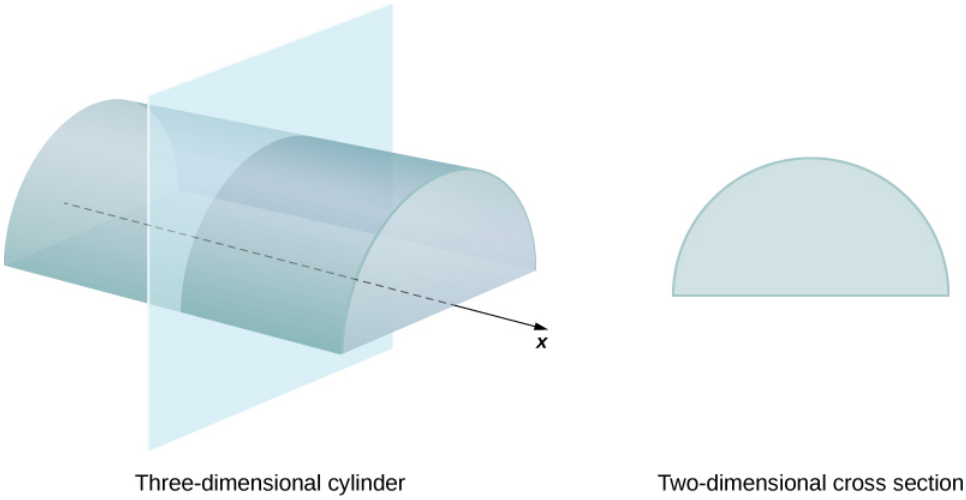


Figure 2.2.1: *Each cross-section of a particular cylinder is identical to the others.*

If a solid does not have a constant cross-section (and it is not one of the other basic solids), we may not have a formula for its volume. In this case, we can use a definite integral to calculate the volume of the solid. We do this by slicing the solid into pieces, estimating the volume of each slice, and then adding those estimated volumes together. The slices should all be parallel to one another, and when we put all the slices together, we should get the whole solid. Consider, for example, the solid S shown in Figure 2.2.2, extending along the x-**axis.**

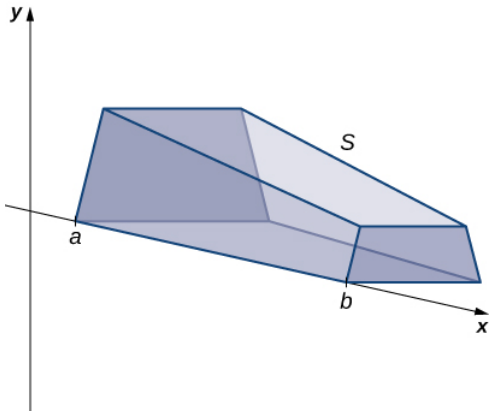


Figure 2.2.2: A solid with a varying cross-section.

We want to divide S into slices perpendicular to the x-**axis**. As we see later in the chapter, there may be times when we want to slice the solid in some other direction—say, with slices perpendicular to the y-axis. The decision of which way to slice the solid is very important. If we make the wrong choice, the computations can get quite messy. Later in the chapter, we examine some of these situations in detail and look at how to decide which way to slice the solid. For the purposes of this section, however, we use slices perpendicular to the x-**axis.**

Because the cross-sectional area is not constant, we let $A(x)$ represent the area of the cross-section at point x. Now let $P = x_0, x_1 \ldots, X_n$ be a regular partition of $[a, b]$, and for $i = 1, 2, \ldots n$, let S_i represent the slice of S stretching from x_{i-1} to x_i. The following figure shows the sliced solid with $n = 3$.

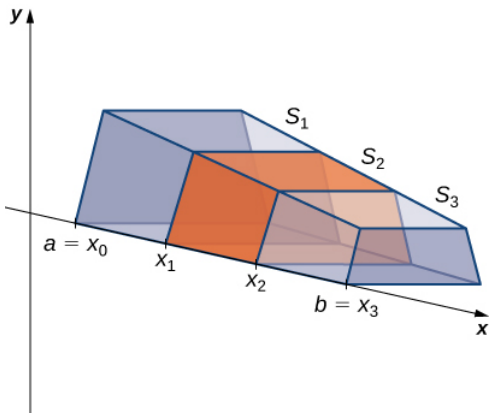


Figure 2.2.3: The solid S has been divided into three slices perpendicular to the x-axis.

Finally, for $i = 1, 2, \ldots n$, let x_i^* be an arbitrary point in $[x_{i-1}, x_i]$. Then the volume of slice S_i can be estimated by $V(S_i) \approx A(x_i^*)\,\Delta x$. Adding these approximations together, we see the volume of the entire solid S can be approximated by

$$V(S) \approx \sum_{i=1}^{n} A(x_i^*)\,\Delta x. \tag{2.2.4}$$

By now, we can recognize this as a Riemann sum, and our next step is to take the limit as $n \to \infty$. Then we have

$$V(S) = \lim_{n\to\infty} \sum_{i=1}^{n} A(x_i^*)\,\Delta x = \int_a^b A(x)\,dx. \tag{2.2.5}$$

The technique we have just described is called the slicing method. To apply it, we use the following strategy.

Problem-Solving Strategy: Finding Volumes by the Slicing Method

1. Examine the solid and determine the shape of a cross-section of the solid. It is often helpful to draw a picture if one is not provided.
2. Determine a formula for the area of the cross-section.
3. Integrate the area formula over the appropriate interval to get the volume.

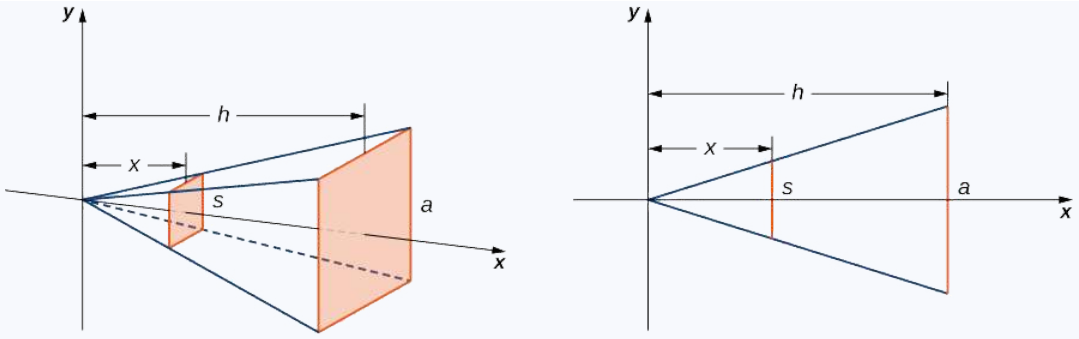
Recall that in this section, we assume the slices are perpendicular to the x-**axis**. Therefore, the area formula is in terms of x and the limits of integration lie on the x-**axis.** However, the problem-solving strategy shown here is valid regardless of how we choose to slice the solid.

Example 2.2.1: Deriving the Formula for the Volume of a Pyramid

We know from geometry that the formula for the volume of a pyramid is $V = \frac{1}{3}Ah$. If the pyramid has a square base, this becomes $V = \frac{1}{3}a^2h$, where a denotes the length of one side of the base. We are going to use the slicing method to derive this formula.

Solution

We want to apply the slicing method to a pyramid with a square base. To set up the integral, consider the pyramid shown in Figure 2.2.4, oriented along the x-**axis.**



(a) (b)

Figure 2.2.4: (a) A pyramid with a square base is oriented along the x-axis. (b) A two-dimensional view of the pyramid is seen from the side.

We first want to determine the shape of a cross-section of the pyramid. We know the base is a square, so the cross-sections are squares as well (step 1). Now we want to determine a formula for the area of one of these cross-sectional squares. Looking at Figure 2.2.4 (b), and using a proportion, since these are similar triangles, we have

$$\frac{s}{a} = \frac{x}{h} \tag{2.2.6}$$

or

$$s = \frac{ax}{h}. \tag{2.2.7}$$

Therefore, the area of one of the cross-sectional squares is

$$A(x) = s^2 = \left(\frac{ax}{h}\right)^2 \qquad \text{(step 2)} \tag{2.2.8}$$

Then we find the volume of the pyramid by integrating from 0 to h (step 3):

$$V = \int_0^h A(x)\,dx = \int_0^h \left(\frac{ax}{h}\right)^2 dx = \frac{a^2}{h^2}\int_0^h x^2\,dx = \left[\frac{a^2}{h^2}\left(\frac{1}{3}x^3\right)\right]\Bigg|_0^h = \frac{1}{3}a^2h. \tag{2.2.9}$$

This is the formula we were looking for.

Exercise 2.2.1

Use the slicing method to derive the formula

$$V = \frac{1}{3}\pi r^2 h \tag{2.2.10}$$

for the volume of a circular cone.

Hint

Use similar triangles, as in Example 2.2.1

Solids of Revolution

If a region in a plane is revolved around a line in that plane, the resulting solid is called a **solid of revolution**, as shown in the following figure.

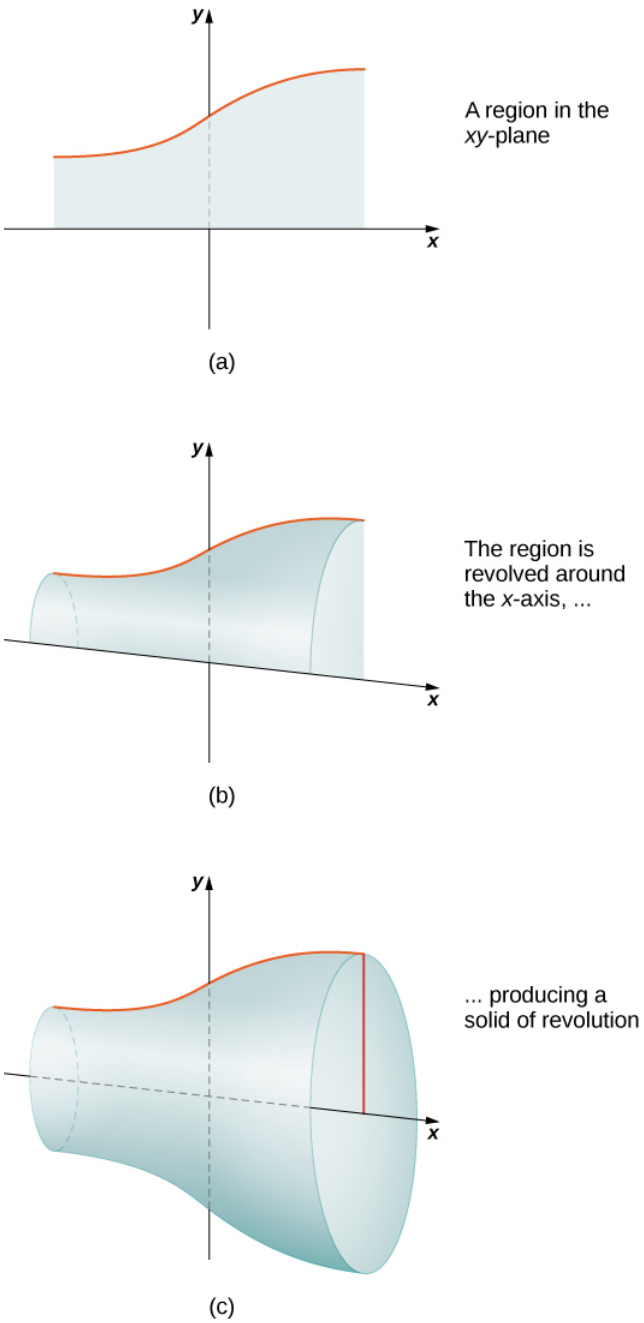


Figure 2.2.5: (a) This is the region that is revolved around the x-axis. (b) As the region begins to revolve around the axis, it sweeps out a solid of revolution. (c) This is the solid that results when the revolution is complete.

Solids of revolution are common in mechanical applications, such as machine parts produced by a lathe. We spend the rest of this section looking at solids of this type. The next example uses the slicing method to calculate the volume of a solid of revolution.

Example 2.2.2: Using the Slicing Method to find the Volume of a Solid of Revolution

Use the slicing method to find the volume of the solid of revolution bounded by the graphs of $f(x) = x^2 - 4x + 5, x = 1$,and $x = 4$, and rotated about the x-axis.

Solution

Using the problem-solving strategy, we first sketch the graph of the quadratic function over the interval $[1, 4]$ as shown in the following figure.

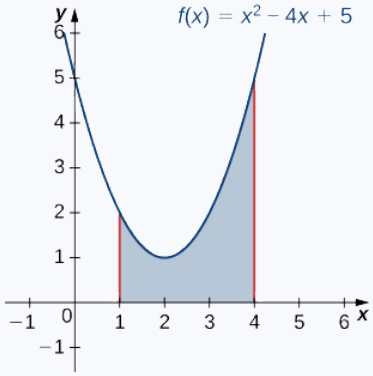
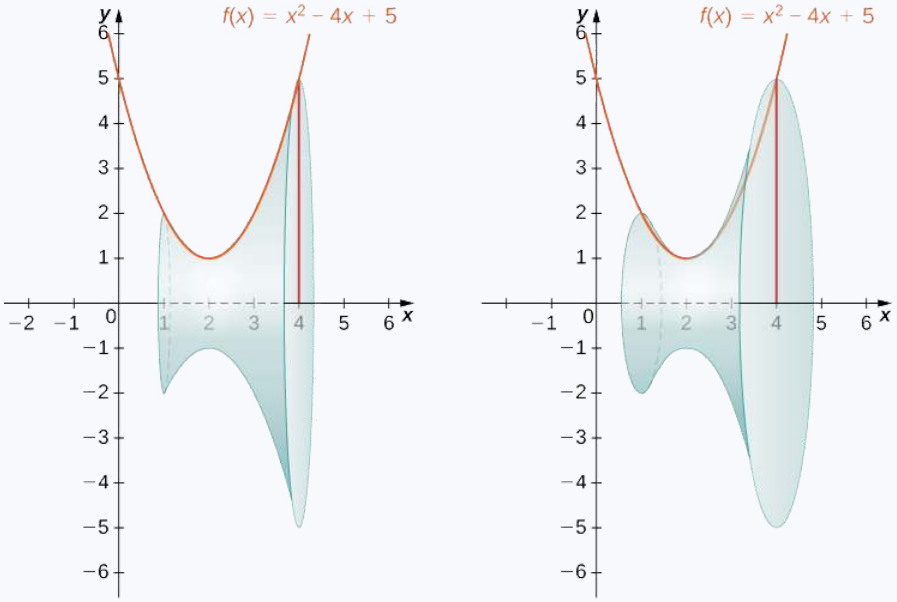


Figure 2.2.6: A region used to produce a solid of revolution.

Next, revolve the region around the x-axis, as shown in the following figure.



(a) (b)

Figure 2.2.7: Two views, (a) and (b), of the solid of revolution produced by revolving the region in Figure 2.2.6 about the x-**axis.**

Since the solid was formed by revolving the region around the x-**axis,** the cross-sections are circles (step 1). The area of the cross-section, then, is the area of a circle, and the radius of the circle is given by $f(x)$. Use the formula for the area of the circle:

$$A(x) = \pi r^2 = \pi[f(x)]^2 = \pi(x^2 - 4x + 5)^2 \qquad \text{(step 2)}. \tag{2.2.11}$$

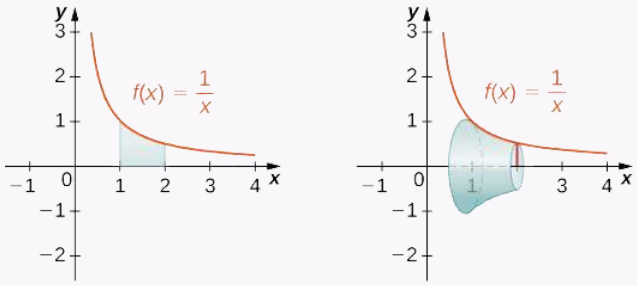
The volume, then, is (step 3)

$$\begin{aligned} V &= \int_a^b A(x)\,dx \\ &= \int_1^4 \pi(x^2 - 4x + 5)^2\,dx \\ &= \pi \int_1^4 (x^4 - 8x^3 + 26x^2 - 40x + 25)\,dx \\ &= \pi \left(\frac{x^5}{5} - 2x^4 + \frac{26x^3}{3} - 20x^2 + 25x \right) \Bigg|_1^4 \\ &= \frac{78}{5}\pi \end{aligned}$$

The volume is $78\pi/5$ units^3.

Exercise 2.2.2

Use the method of slicing to find the volume of the solid of revolution formed by revolving the region between the graph of the function $f(x) = 1/x$ and the x-**axis** over the interval $[1, 2]$ around the x-**axis.** See the following figure.



(a) (b)

Hint

Use the problem-solving strategy presented earlier and follow Example 2.2.2to help with step 2.

Answer

$\frac{\pi}{2}$ units^3

The Disk Method

When we use the slicing method with solids of revolution, it is often called the disk method because, for solids of revolution, the slices used to over approximate the volume of the solid are disks. To see this, consider the solid of revolution generated by revolving the region between the graph of the function $f(x) = (x-1)^2 + 1$ and the x-**axis** over the interval $[-1, 3]$ around the x-**axis**. The graph of the function and a representative disk are shown in Figure 2.2.8 (a) and (b). The region of revolution and the resulting solid are shown in Figure 2.2.8 (c) and (d).

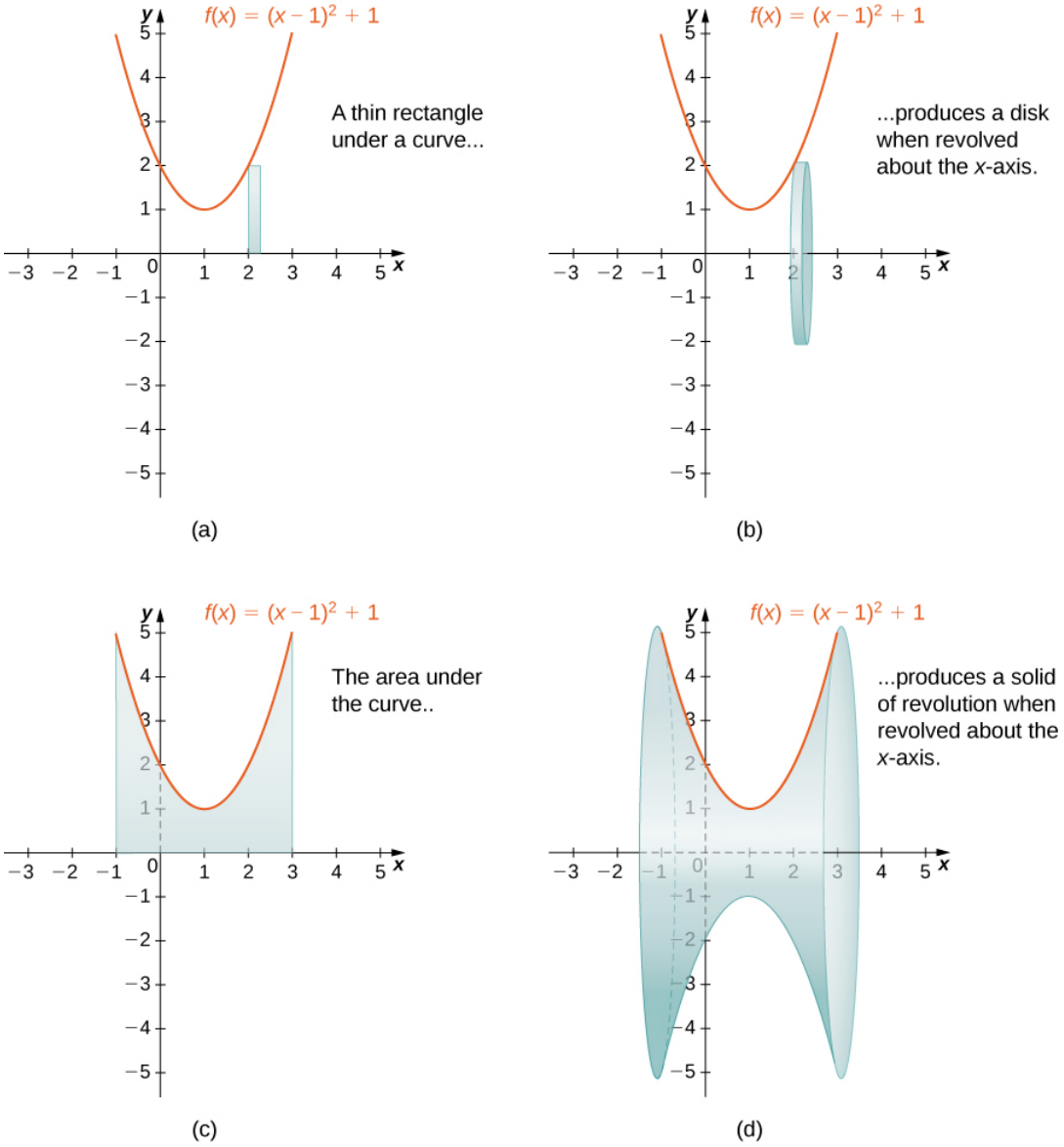


Figure 2.2.8: (a) A thin rectangle for approximating the area under a curve. (b) A representative disk formed by revolving the rectangle about the x**-axis.** (c) The region under the curve is revolved about the x**-axis,** resulting in (d) the solid of revolution.

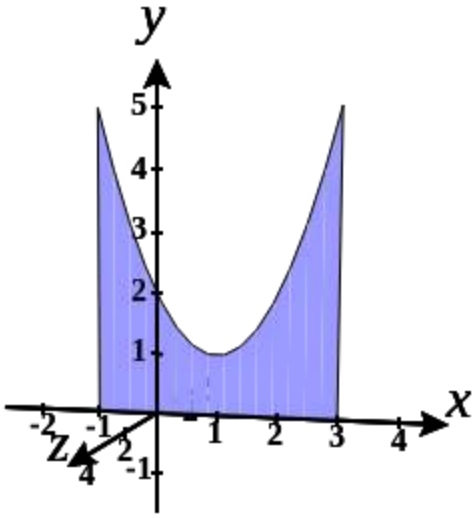


Figure 2.2.8: (e) A dynamic version of this solid of revolution generated using CalcPlot3D.

We already used the formal Riemann sum development of the volume formula when we developed the slicing method. We know that

$$\int_a^b A(x)\,dx.$$

The only difference with the disk method is that we know the formula for the cross-sectional area ahead of time; it is the area of a circle. This gives the following rule.

The Disk Method

Let $f(x)$ be continuous and nonnegative. Define R as the region bounded above by the graph of $f(x)$, below by the x-**axis**, on the left by the line $x = a$, and on the right by the line $x = b$. Then, the volume of the solid of revolution formed by revolving R around the x-**axis** is given by

$$V = \int_a^b \pi[f(x)]^2\,dx. \tag{2.2.12}$$

The volume of the solid we have been studying (Figure 2.2.8) is given by

$$
\begin{aligned}
V &= \int_a^b \pi [f(x)]^2 \, dx \\
&= \int_{-1}^{3} \pi \left[(x-1)^2 + 1 \right]^2 dx = \pi \int_{-1}^{3} \left[(x-1)^4 + 2(x-1)^2 + 1 \right]^2 dx \\
&= \pi \left[\frac{1}{5}(x-1)^5 + \frac{2}{3}(x-1)^3 + x \right] \Big|_{-1}^{3} \\
&= \pi \left[\left(\frac{32}{5} + \frac{16}{3} + 3 \right) - \left(-\frac{32}{5} - \frac{16}{3} - 1 \right) \right] \\
&= \frac{412\pi}{15} \text{ units}^3.
\end{aligned}
$$

Let's look at some examples.

Example 2.2.3: Using the Disk Method to Find the Volume of a Solid of Revolution 1

Use the disk method to find the volume of the solid of revolution generated by rotating the region between the graph of $f(x) = \sqrt{x}$ and the x-**axis** over the interval $[1, 4]$ around the x-**axis.**

Solution

The graphs of the function and the solid of revolution are shown in the following figure.

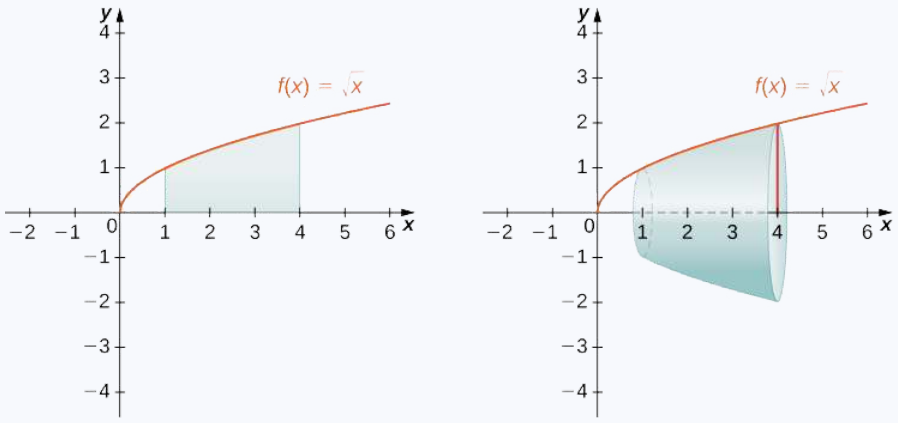


Figure 2.2.9: (a) The function $f(x) = \sqrt{x}$ over the interval $[1, 4]$. (b) The solid of revolution obtained by revolving the region under the graph of $f(x)$ about the x-**axis**.

We have

$$
\begin{aligned}
V &= \int_a^b \pi \left[f(x) \right]^2 dx \\
&= \int_1^4 \pi [\sqrt{x}]^2 \, dx = \pi \int_1^4 x \, dx \\
&= \frac{\pi}{2} x^2 \Big|_1^4 = \frac{15\pi}{2}
\end{aligned}
$$

The volume is $(15\pi)/2 \text{ units}^3$.

Exercise 2.2.3

Use the disk method to find the volume of the solid of revolution generated by rotating the region between the graph of $f(x) = \sqrt{4 - x}$ and the x-**axis** over the interval $[0, 4]$ around the x-**axis.**

Hint

Use the procedure from Example 2.2.3

Answer

8π units^3

So far, our examples have all concerned regions revolved around the x-**axis**, but we can generate a solid of revolution by revolving a plane region around any horizontal or vertical line. In the next example, we look at a solid of revolution that has been generated by revolving a region around the y-**axis**. The mechanics of the disk method are nearly the same as when the x-**axis** is the axis of revolution, but we express the function in terms of y and we integrate with respect to y as well. This is summarized in the following rule.

Rule: The Disk Method for Solids of Revolution around the y-axis

Let $g(y)$ be continuous and nonnegative. Define Q as the region bounded on the right by the graph of $g(y)$, on the left by the y-**axis**, below by the line $y = c$, and above by the line $y = d$. Then, the volume of the solid of revolution formed by revolving Q around the y-**axis** is given by

$$V = \int_c^d \pi \left[g(y)\right]^2 dy. \tag{2.2.13}$$

The next example shows how this rule works in practice.

Example 2.2.4: Using the Disk Method to Find the Volume of a Solid of Revolution 2

Let R be the region bounded by the graph of $g(y) = \sqrt{4 - y}$ and the y-**axis** over the y-**axis** interval $[0, 4]$. Use the disk method to find the volume of the solid of revolution generated by rotating R around the y-**axis.**

Solution

Figure 2.2.10 shows the function and a representative disk that can be used to estimate the volume. Notice that since we are revolving the function around the y-**axis,** the disks are horizontal, rather than vertical.

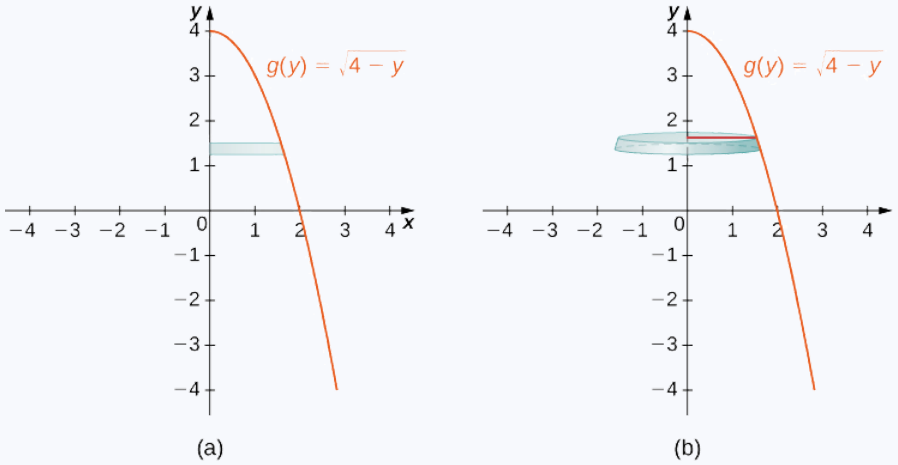


Figure 2.2.10: (a) Shown is a thin rectangle between the curve of the function $g(y) = \sqrt{4 - y}$ and the y-**axis.** (b) The rectangle forms a representative disk after revolution around the y-**axis.**

The region to be revolved and the full solid of revolution are depicted in the following figure.

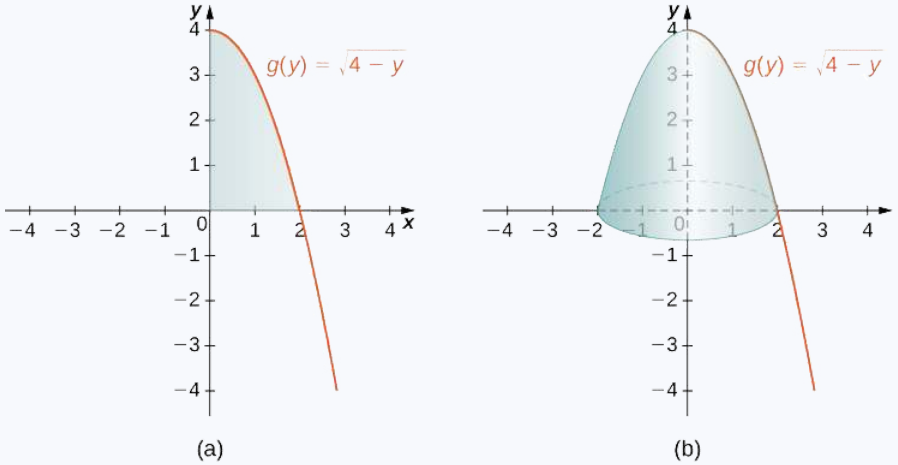


Figure 2.2.11: (a) The region to the left of the function $g(y) = \sqrt{4-y}$ over the y-axis interval $[0, 4]$. (b) The solid of revolution formed by revolving the region about the y-**axis.**

To find the volume, we integrate with respect to y We obtain

$$V = \int_c^d \pi \left[g(y)\right]^2 dy = \int_0^4 \pi \left[\sqrt{4-y}\right]^2 dy = \pi \int_0^4 (4-y)\, dy = \pi \left[4y - \frac{y^2}{2}\right]\Big|_0^4 = 8\pi. \tag{2.2.14}$$

The volume is 8π units^3.

Exercise 2.2.4

Use the disk method to find the volume of the solid of revolution generated by rotating the region between the graph of $g(y) = y$ and the y-**axis** over the interval $[1, 4]$ around the y-**axis.**

Hint

Use the procedure from Example 2.2.4

Answer

21π units^3

The Washer Method

Some solids of revolution have cavities in the middle; they are not solid all the way to the axis of revolution. Sometimes, this is just a result of the way the region of revolution is shaped with respect to the axis of revolution. In other cases, cavities arise when the region of revolution is defined as the region between the graphs of two functions. A third way this can happen is when an axis of revolution other than the x-**axis** or y-**axis** is selected.

When the solid of revolution has a cavity in the middle, the slices used to approximate the volume are not disks, but washers (disks with holes in the center). For example, consider the region bounded above by the graph of the function $f(x) = \sqrt{x}$ and below by the graph of the function $g(x) = 1$ over the interval $[1, 4]$. When this region is revolved around the x-**axis**, the result is a solid with a cavity in the middle, and the slices are washers. The graph of the function and a representative washer are shown in Figure 2.2.12 (a) and (b). The region of revolution and the resulting solid are shown in Figure 2.2.12(c) and (d).

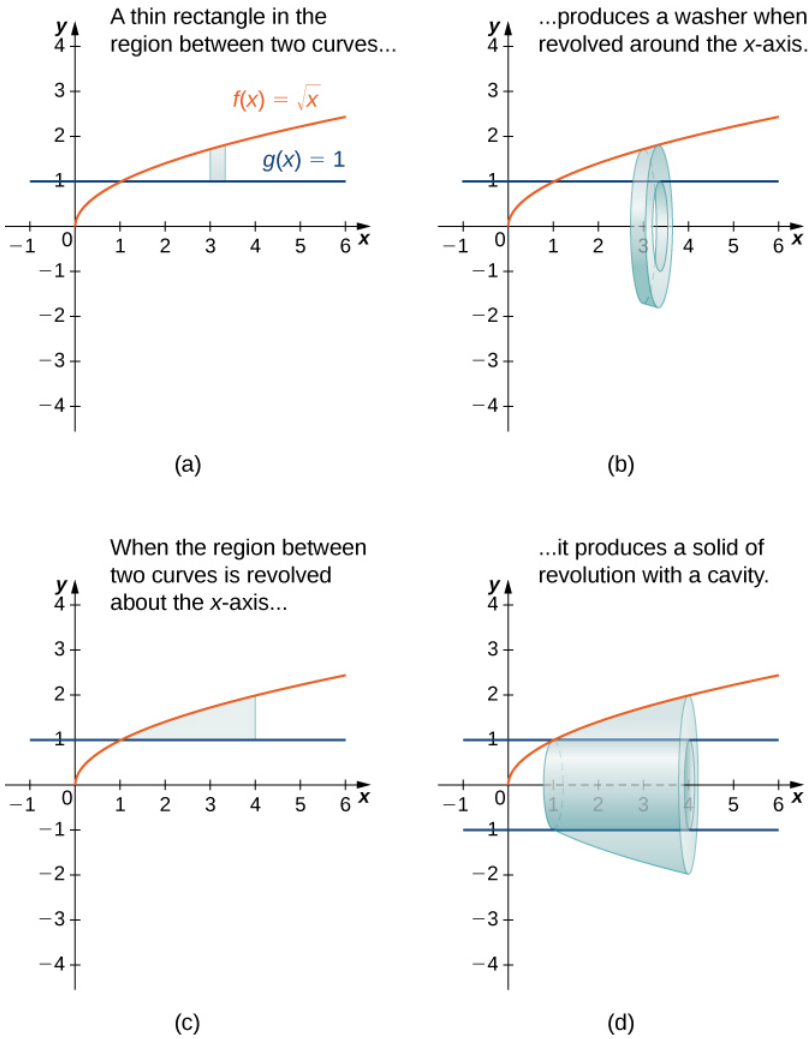


Figure 2.2.12: (a) A thin rectangle in the region between two curves. (b) A representative disk formed by revolving the rectangle about the x-**axis.** (c) The region between the curves over the given interval. (d) The resulting solid of revolution.

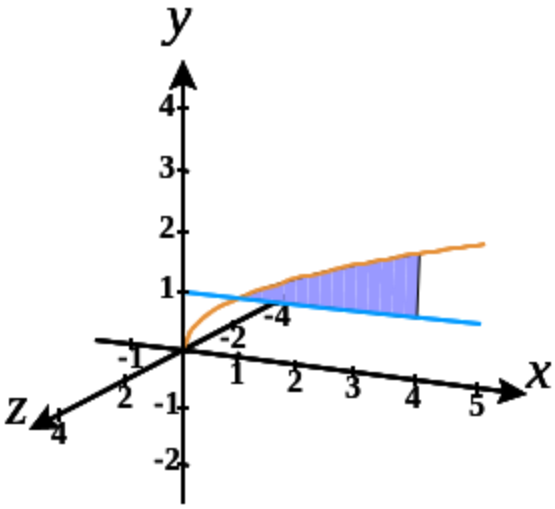


Figure 2.2.12: (e) A dynamic version of this solid of revolution generated using CalcPlot3D.

The cross-sectional area, then, is the area of the outer circle less the area of the inner circle. In this case,

$$A(x) = \pi(\sqrt{x})^2 - \pi(1)^2 = \pi(x-1).$$

Then the volume of the solid is

$$V = \int_a^b A(x)\,dx = \int_1^4 \pi(x-1)\,dx = \pi\left[\frac{x^2}{2} - x\right]\Big|_1^4 = \frac{9}{2}\pi \text{ units}^3. \tag{2.2.15}$$

Generalizing this process gives the **washer method.**

Rule: The Washer Method

Suppose $f(x)$ and $g(x)$ are continuous, nonnegative functions such that $f(x) \geq g(x)$ over $[a,b]$. Let R denote the region bounded above by the graph of $f(x)$, below by the graph of $g(x)$, on the left by the line $x = a$, and on the right by the line $x = b$. Then, the volume of the solid of revolution formed by revolving R around the x**-axis** is given by

$$V = \int_a^b \pi\left[(f(x))^2 - (g(x))^2\right]\,dx. \tag{2.2.16}$$

Example 2.2.5: Using the Washer Method

Find the volume of a solid of revolution formed by revolving the region bounded above by the graph of $f(x) = x$ and below by the graph of $g(x) = 1/x$ over the interval $[1,4]$ around the x**-axis.**

Solution

The graphs of the functions and the solid of revolution are shown in the following figure.

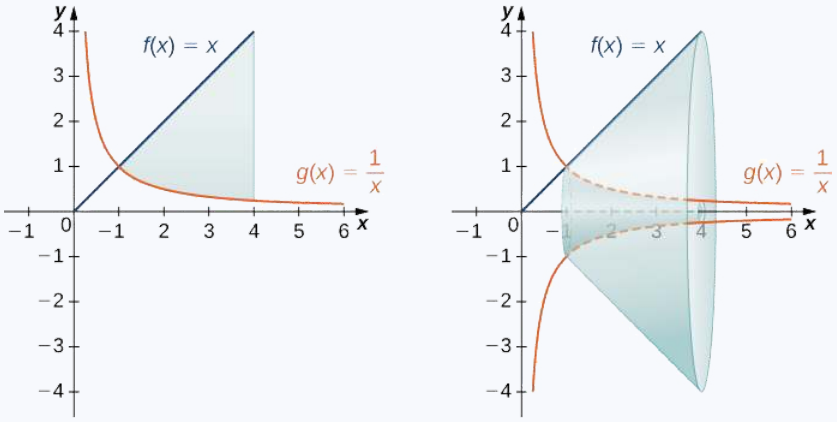


Figure 2.2.13: (a) The region between the graphs of the functions $f(x) = x$ and $g(x) = 1/x$ over the interval $[1, 4]$. (b) Revolving the region about the x-**axis** generates a solid of revolution with a cavity in the middle.

We have

$$\begin{aligned} V &= \int_a^b \pi\left[(f(x))^2 - (g(x))^2\right] dx = \pi \int_1^4 \left[x^2 - \left(\frac{1}{x}\right)^2\right] dx \\ &= \pi \left[\frac{x^3}{3} + \frac{1}{x}\right]\Bigg|_1^4 \\ &= \frac{81\pi}{4} \text{ units}^3. \end{aligned}$$

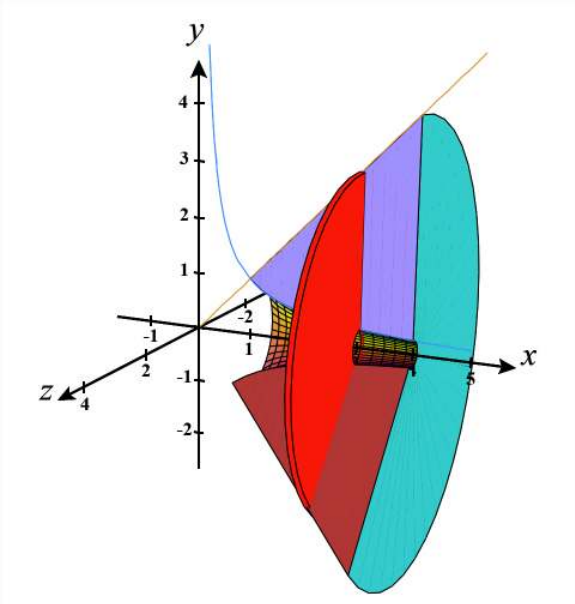


Figure 2.2.13: (c) A dynamic version of this solid of revolution generated using CalcPlot3D.

Exercise 2.2.5

Find the volume of a solid of revolution formed by revolving the region bounded by the graphs of $f(x) = \sqrt{x}$ and $g(x) = 1/x$ over the interval $[1, 3]$ around the x**-axis.**

Hint

Graph the functions to determine which graph forms the upper bound and which graph forms the lower bound, then use the procedure from Example 2.2.5

Answer

$\frac{10\pi}{3}$ units^3

As with the disk method, we can also apply the washer method to solids of revolution that result from revolving a region around the y-axis. In this case, the following rule applies.

Rule: The Washer Method for Solids of Revolution around the y-axis

Suppose $u(y)$ and $v(y)$ are continuous, nonnegative functions such that $v(y) \leq u(y)$ for $y \in [c, d]$. Let Q denote the region bounded on the right by the graph of $u(y)$, on the left by the graph of $v(y)$, below by the line $y = c$, and above by the line $y = d$. Then, the volume of the solid of revolution formed by revolving Q around the y-**axis** is given by

$$V = \int_c^d \pi \left[(u(y))^2 - (v(y))^2 \right] \, dy. \tag{2.2.17}$$

Rather than looking at an example of the washer method with the y-**axis** as the axis of revolution, we now consider an example in which the axis of revolution is a line other than one of the two coordinate axes. The same general method applies, but you may have to visualize just how to describe the cross-sectional area of the volume.

Example 2.2.6:

Find the volume of a solid of revolution formed by revolving the region bounded above by $f(x) = 4 - x$ and below by the x-**axis** over the interval $[0, 4]$ around the line $y = -2$.

Solution

The graph of the region and the solid of revolution are shown in the following figure.

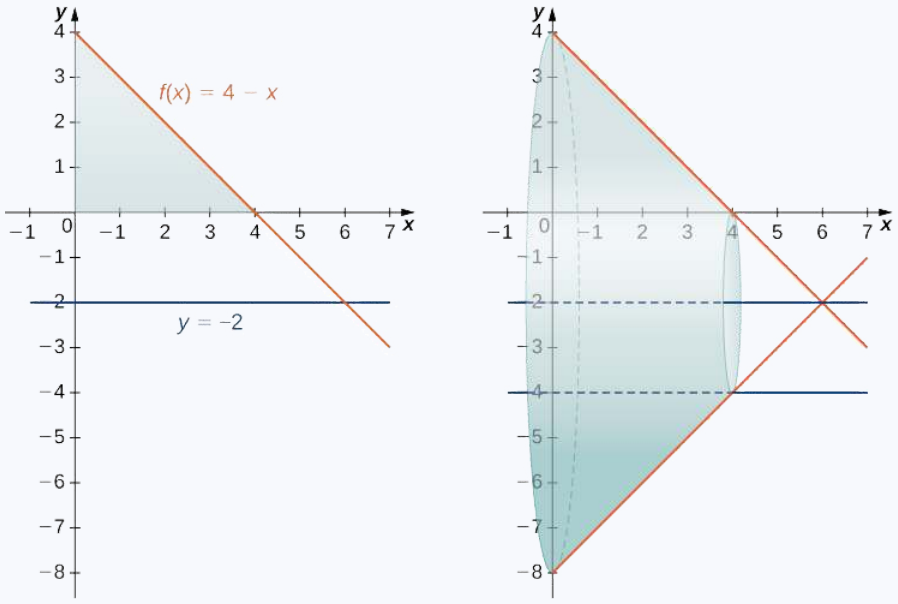


Figure 2.2.14: (a) The region between the graph of the function $f(x) = 4 - x$ and the x-**axis** over the interval $[0, 4]$. (b) Revolving the region about the line $y = -2$ generates a solid of revolution with a cylindrical hole through its middle.

We can't apply the volume formula to this problem directly because the axis of revolution is not one of the coordinate axes. However, we still know that the area of the cross-section is the area of the outer circle less the area of the inner circle. Looking at the graph of the function, we see the radius of the outer circle is given by $f(x) + 2$, which simplifies to

$$f(x)+2=(4-x)+2=6-x.$$

The radius of the inner circle is $g(x)=2$. Therefore, we have

$$\begin{aligned} V &= \int_0^4 \pi\left[(6-x)^2-(2)^2\right]\,dx \\ &= \pi\int_0^4 (x^2-12x+32)\,dx = \pi\left[\frac{x^3}{3}-6x^2+32x\right]\Big|_0^4 \\ &= \frac{160\pi}{3}\,\text{units}^3. \end{aligned}$$

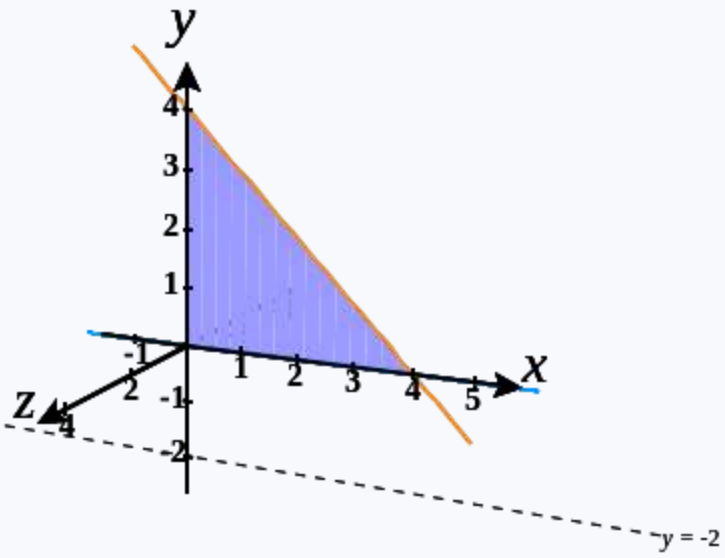


Figure 2.2.14: (c) A dynamic version of this solid of revolution generated using CalcPlot3D.

Exercise 2.2.6

Find the volume of a solid of revolution formed by revolving the region bounded above by the graph of $f(x)=x+2$ and below by the x-**axis** over the interval $[0,3]$ around the line $y=-1$.

Hint

Use the procedure from Example 2.2.6

Answer

$60\pi\,\text{units}^3$

Key Concepts

- Definite integrals can be used to find the volumes of solids. Using the slicing method, we can find a volume by integrating the cross-sectional area.
- For solids of revolution, the volume slices are often disks and the cross-sections are circles. The method of disks involves applying the method of slicing in the particular case in which the cross-sections are circles, and using the formula for the area of a circle.
- If a solid of revolution has a cavity in the center, the volume slices are washers. With the method of washers, the area of the inner circle is subtracted from the area of the outer circle before integrating.

Key Equations

- **Disk Method along the x-axis**

$$V = \int_a^b \pi \big[f(x)\big]^2 \, dx$$

- **Disk Method along the y-axis**

$$V = \int_c^d \pi \big[g(y)\big]^2 \, dy$$

- **Washer Method**

$$V = \int_a^b \pi \left[(f(x))^2 - (g(x))^2\right] \, dx$$

Glossary

cross-section
the intersection of a plane and a solid object

disk method
a special case of the slicing method used with solids of revolution when the slices are disks

slicing method
a method of calculating the volume of a solid that involves cutting the solid into pieces, estimating the volume of each piece, then adding these estimates to arrive at an estimate of the total volume; as the number of slices goes to infinity, this estimate becomes an integral that gives the exact value of the volume

solid of revolution
a solid generated by revolving a region in a plane around a line in that plane

washer method
a special case of the slicing method used with solids of revolution when the slices are washers

Contributors

- Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 2.2: Determining Volumes by Slicing is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Volumes of Revolution - Cylindrical Shells

Learning Objectives

- Calculate the volume of a solid of revolution by using the method of cylindrical shells.
- Compare the different methods for calculating a volume of revolution.

In this section, we examine the method of cylindrical shells, the final method for finding the volume of a solid of revolution. We can use this method on the same kinds of solids as the disk method or the washer method; however, with the disk and washer methods, we integrate along the coordinate axis parallel to the axis of revolution. With the method of cylindrical shells, we integrate along the coordinate *axis perpendicular* to the axis of revolution. The ability to choose which variable of integration we want to use can be a significant advantage with more complicated functions. Also, the specific geometry of the solid sometimes makes the method of using cylindrical shells more appealing than using the washer method. In the last part of this section, we review all the methods for finding volume that we have studied and lay out some guidelines to help you determine which method to use in a given situation.

The Method of Cylindrical Shells

Again, we are working with a solid of revolution. As before, we define a region R, bounded above by the graph of a function $y = f(x)$, below by the $\boldsymbol{x}$**-axis**, and on the left and right by the lines $x = a$ and $x = b$, respectively, as shown in Figure $2.3.1a$. We then revolve this region around the y-axis, as shown in Figure $2.3.1b$. Note that this is different from what we have done before. Previously, regions defined in terms of functions of x were revolved around the $\boldsymbol{x}$**-axis** or a line parallel to it.

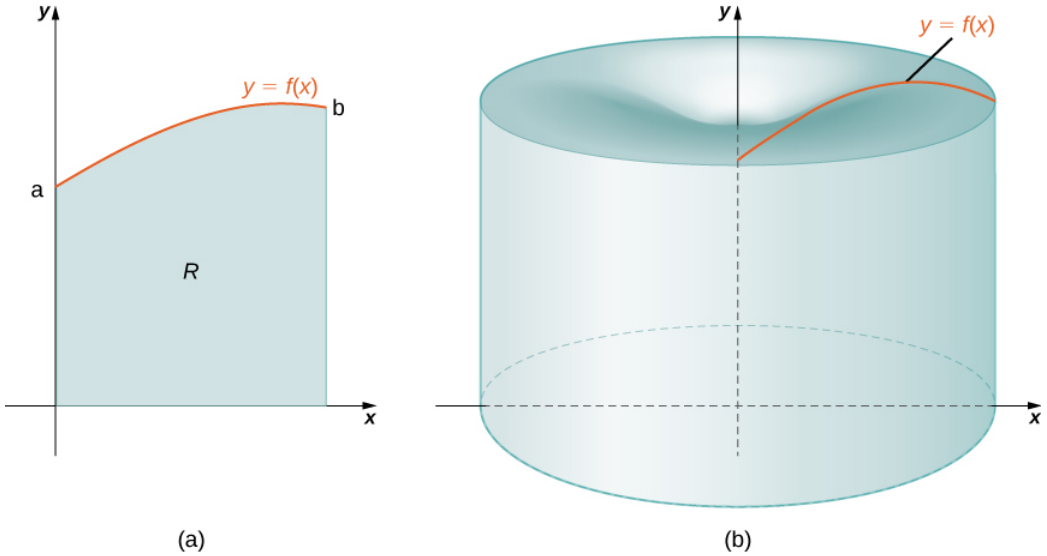


Figure 2.3.1: (a) A region bounded by the graph of a function of x. (b) The solid of revolution formed when the region is revolved around the $\boldsymbol{y}$**-axis.**

As we have done many times before, partition the interval $[a, b]$ using a regular partition, $P = x_0, x_1, \ldots, x_n$ and, for $i = 1, 2, \ldots, n$, choose a point $x_i^* \in [x_{i-1}, x_i]$. Then, construct a rectangle over the interval $[x_{i-1}, x_i]$ of height $f(x_i^*)$ and width Δx. A representative rectangle is shown in Figure $2.3.2a$. When that rectangle is revolved around the y-axis, instead of a disk or a washer, we get a cylindrical shell, as shown in Figure $2.3.2$.

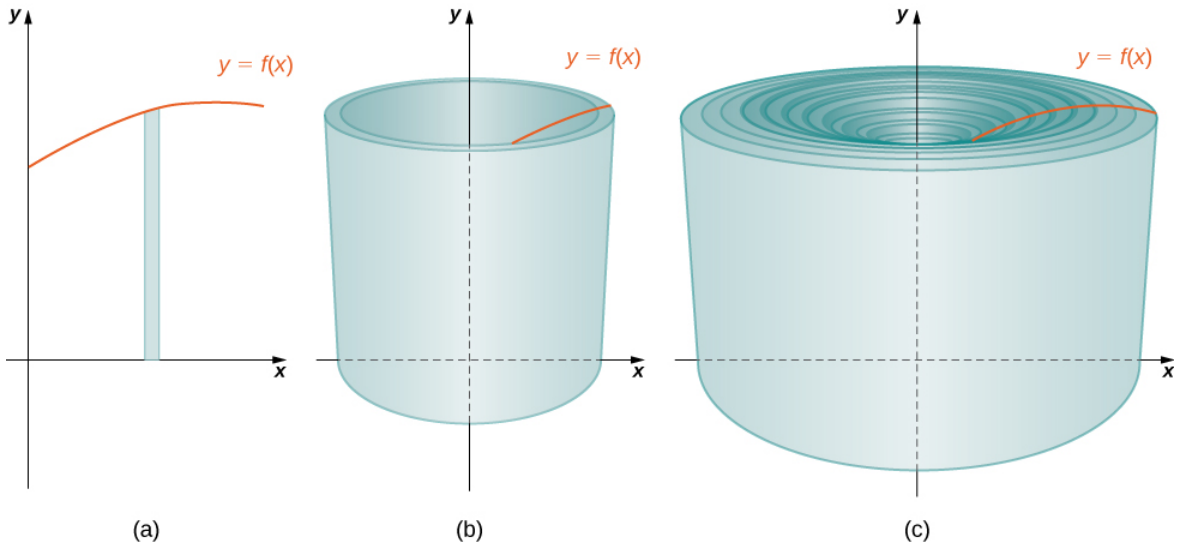


Figure 2.3.2: (a) A representative rectangle. (b) When this rectangle is revolved around the ***y*-axis**, the result is a cylindrical shell. (c) When we put all the shells together, we get an approximation of the original solid.

To calculate the volume of this shell, consider Figure 2.3.3.

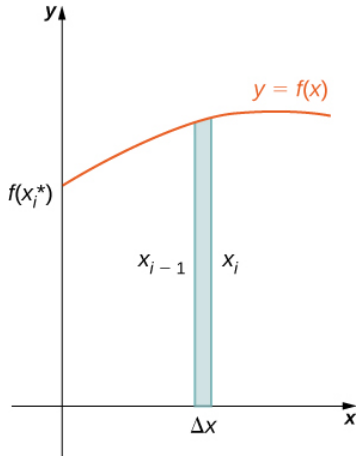


Figure 2.3.3: Calculating the volume of the shell.

The shell is a cylinder, so its volume is the cross-sectional area multiplied by the height of the cylinder. The cross-sections are annuli (ring-shaped regions—essentially, circles with a hole in the center), with outer radius x_i and inner radius x_{i-1}. Thus, the cross-sectional area is $\pi x_i^2 - \pi x_{i-1}^2$. The height of the cylinder is $f(x_i^*)$. Then the volume of the shell is

$$\begin{aligned} V_{shell} &= f(x_i^*)(\pi\, x_i^2 - \pi\, x_{i-1}^2) \\ &= \pi\, f(x_i^*)(x_i^2 - x_{i-1}^2) \\ &= \pi\, f(x_i^*)(x_i + x_{i-1})(x_i - x_{i-1}) \\ &= 2\pi\, f(x_i^*)\left(\frac{x_i + x_{i-1}}{2}\right)(x_i - x_{i-1}). \end{aligned}$$

Note that $x_i - x_{i-1} = \Delta x,$ so we have

$$V_{shell} = 2\pi\, f(x_i^*)\left(\frac{x_i + x_{i-1}}{2}\right)\Delta x.$$

Furthermore, $\dfrac{x_i + x_{i-1}}{2}$ is both the midpoint of the interval $[x_{i-1}, x_i]$ and the average radius of the shell, and we can approximate this by x_i^*. We then have

$$V_{shell} \approx 2\pi\, f(x_i^*)x_i^*\, \Delta x.$$

Another way to think of this is to think of making a vertical cut in the shell and then opening it up to form a flat plate (Figure 2.3.4).

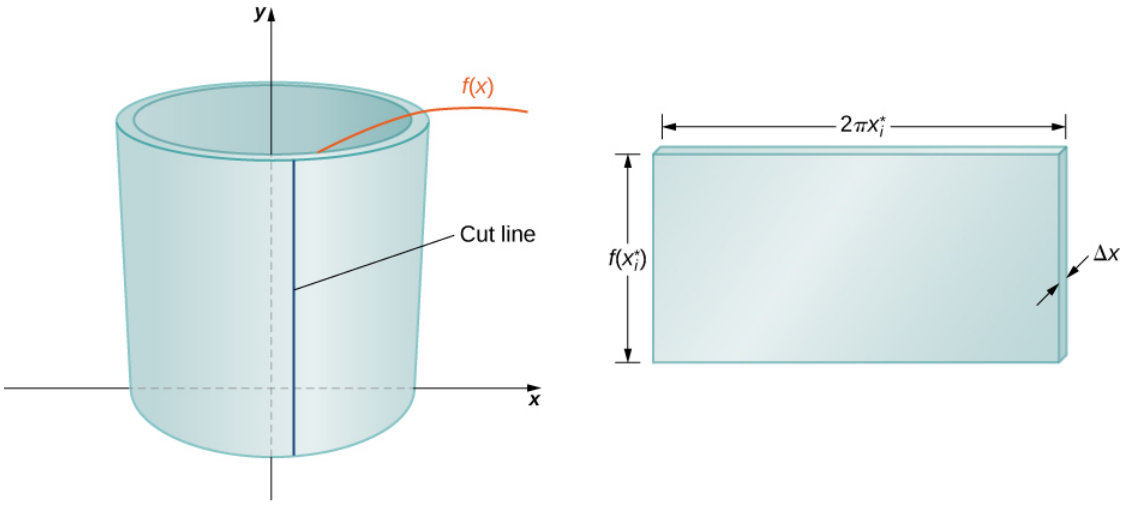


Figure 2.3.4: (a) Make a vertical cut in a representative shell. (b) Open the shell up to form a flat plate.

In reality, the outer radius of the shell is greater than the inner radius, and hence the back edge of the plate would be slightly longer than the front edge of the plate. However, we can approximate the flattened shell by a flat plate of height $f(x_i^*)$, width $2\pi x_i^*$, and thickness Δx (Figure). The volume of the shell, then, is approximately the volume of the flat plate. Multiplying the height, width, and depth of the plate, we get

$$V_{shell} \approx f(x_i^*)(2\pi\, x_i^*)\,\Delta x,$$

which is the same formula we had before.

To calculate the volume of the entire solid, we then add the volumes of all the shells and obtain

$$V \approx \sum_{i=1}^{n} (2\pi\, x_i^* f(x_i^*)\,\Delta x).$$

Here we have another Riemann sum, this time for the function $2\pi\, x\, f(x)$.Taking the limit as $n \to \infty$ gives us

$$V = \lim_{n\to\infty} \sum_{i=1}^{n} (2\pi\, x_i^* f(x_i^*)\,\Delta x) = \int_a^b (2\pi\, x\, f(x))\, dx.$$

This leads to the following rule for the *method of cylindrical shells*.

Rule: The Method of Cylindrical Shells

Let $f(x)$ be continuous and nonnegative. Define R as the region bounded above by the graph of $f(x)$, below by the **x-axis,** on the left by the line $x = a$, and on the right by the line $x = b$. Then the volume of the solid of revolution formed by revolving R around the y-axis is given by

$$V = \int_a^b (2\pi\, x\, f(x))\, dx.$$

Now let's consider an example.

Example 2.3.1: The Method of Cylindrical Shells I

Define R as the region bounded above by the graph of $f(x) = 1/x$ and below by the **x-axis** over the interval $[1, 3]$. Find the volume of the solid of revolution formed by revolving R around the y-axis.

Solution

First we must graph the region R and the associated solid of revolution, as shown in Figure 2.3.5.

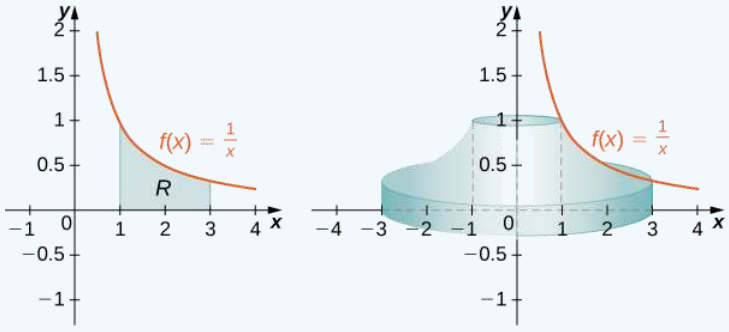


Figure 2.3.5: (a) The region R under the graph of $f(x) = 1/x$ over the interval $[1, 3]$. (b) The solid of revolution generated by revolving R about the y-axis.

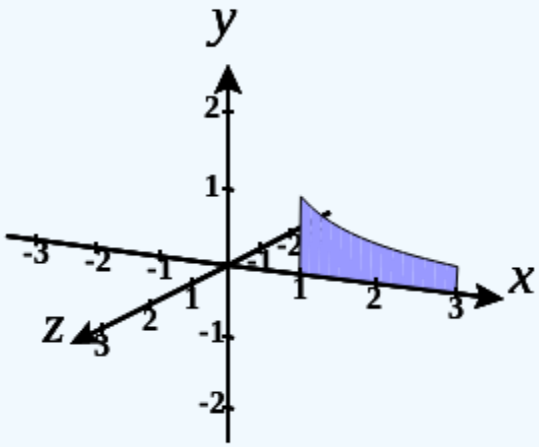


Figure 2.3.5 (c) Visualizing the solid of revolution with CalcPlot3D.

Then the volume of the solid is given by

$$V = \int_a^b (2\pi\, x\, f(x))\, dx$$
$$= \int_1^3 \left(2\pi\, x \left(\frac{1}{x}\right)\right) dx$$
$$= \int_1^3 2\pi\, dx$$
$$= 2\pi\, x\Big|_1^3 = 4\pi \text{ units}^3.$$

Exercise 2.3.1

Define R as the region bounded above by the graph of $f(x) = x^2$ and below by the x-axis over the interval $[1, 2]$. Find the volume of the solid of revolution formed by revolving R around the y-axis.

Hint

Use the procedure from Example 2.3.1

Answer

$\frac{15\pi}{2}$ units^3

Example 2.3.2: The Method of Cylindrical Shells II

Define R as the region bounded above by the graph of $f(x) = 2x - x^2$ and below by the x-axis over the interval $[0, 2]$. Find the volume of the solid of revolution formed by revolving R around the **y-axis.**

Solution

First graph the region R and the associated solid of revolution, as shown in Figure 2.3.6.

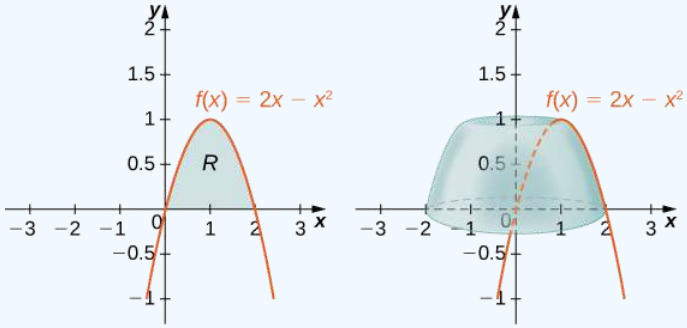


Figure 2.3.6: (a) The region R *under the graph of* $f(x) = 2x - x^2$ *over the interval* $[0, 2]$. *(b) The volume of revolution obtained by revolving* R *about the* y*-axis.*

Then the volume of the solid is given by

$$
\begin{aligned}
V &= \int_a^b (2\pi\, x\, f(x))\, dx \\
&= \int_0^2 (2\pi\, x(2x - x^2))\, dx \\
&= 2\pi \int_0^2 (2x^2 - x^3)\, dx \\
&= 2\pi \left[\frac{2x^3}{3} - \frac{x^4}{4} \right] \Big|_0^2 \\
&= \frac{8\pi}{3} \text{ units}^3
\end{aligned}
$$

Exercise 2.3.2

Define R as the region bounded above by the graph of $f(x) = 3x - x^2$ and below by the x-axis over the interval $[0, 2]$. Find the volume of the solid of revolution formed by revolving R around the y-axis.

Hint

Use the process from Example 2.3.2

Answer

$8\pi \text{ units}^3$

As with the disk method and the washer method, we can use the method of cylindrical shells with solids of revolution, revolved around the x-axis, when we want to integrate with respect to y. The analogous rule for this type of solid is given here.

Rule: The Method of Cylindrical Shells for Solids of Revolution around the x-axis

Let $g(y)$ be continuous and nonnegative. Define Q as the region bounded on the right by the graph of $g(y)$, on the left by the y-axis, below by the line $y = c$, and above by the line $y = d$. Then, the volume of the solid of revolution formed by revolving Q around the x-axis is given by

$$V = \int_c^d (2\pi\, y\, g(y))\, dy.$$

Example 2.3.3: The Method of Cylindrical Shells for a Solid Revolved around the x-axis

Define Q as the region bounded on the right by the graph of $g(y) = 2\sqrt{y}$ and on the left by the y-axis for $y \in [0, 4]$. Find the volume of the solid of revolution formed by revolving Q around the x-axis.

Solution

First, we need to graph the region Q and the associated solid of revolution, as shown in Figure 2.3.7.

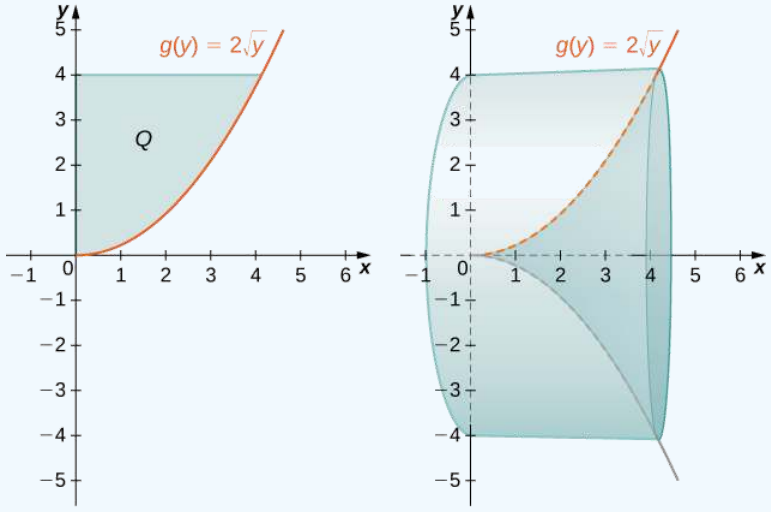


Figure 2.3.7: (a) The region Q to the left of the function $g(y)$ over the interval $[0,4]$. (b) The solid of revolution generated by revolving Q around the x-axis.

Label the shaded region Q. Then the volume of the solid is given by

$$\begin{aligned} V &= \int_c^d (2\pi\, y\, g(y))\, dy \\ &= \int_0^4 (2\pi\, y(2\sqrt{y}))\, dy \\ &= 4\pi \int_0^4 y^{3/2}\, dy \\ &= 4\pi \left[\frac{2y^{5/2}}{5}\right]\Bigg|_0^4 \\ &= \frac{256\pi}{5} \text{ units}^3 \end{aligned}$$

? Exercise 2.3.3

Define Q as the region bounded on the right by the graph of $g(y) = 3/y$ and on the left by the y-axis for $y \in [1,3]$. Find the volume of the solid of revolution formed by revolving Q around the x-axis.

Hint

Use the process from Example 2.3.3

Answer

12π units^3

For the next example, we look at a solid of revolution for which the graph of a function is revolved around a line other than one of the two coordinate axes. To set this up, we need to revisit the development of the method of cylindrical shells. Recall that we found the volume of one of the shells to be given by

$$\begin{aligned} V_{shell} &= f(x_i^*)(\pi\, x_i^2 - \pi\, x_{i-1}^2) \\ &= \pi\, f(x_i^*)(x_i^2 - x_{i-1}^2) \\ &= \pi\, f(x_i^*)(x_i + x_{i-1})(x_i - x_{i-1}) \\ &= 2\pi\, f(x_i^*)\left(\frac{x_i + x_{i-1}}{2}\right)(x_i - x_{i-1}). \end{aligned}$$

This was based on a shell with an outer radius of x_i and an inner radius of x_{i-1}. If, however, we rotate the region around a line other than the y-axis, we have a different outer and inner radius. Suppose, for example, that we rotate the region around the line $x = -k$, where k is some positive constant. Then, the outer radius of the shell is $x_i + k$ and the inner radius of the shell is $x_{i-1} + k$. Substituting these terms into the expression for volume, we see that when a plane region is rotated around the line $x = -k$, the volume of a shell is given by

$$\begin{aligned} V_{shell} &= 2\pi\, f(x_i^*)\left(\frac{(x_i + k) + (x_{i-1} + k)}{2}\right)((x_i + k) - (x_{i-1} + k)) \\ &= 2\pi\, f(x_i^*)\left(\left(\frac{x_i + x_{i-2}}{2}\right) + k\right)\Delta x. \end{aligned}$$

As before, we notice that $\dfrac{x_i + x_{i-1}}{2}$ is the midpoint of the interval $[x_{i-1}, x_i]$ and can be approximated by x_i^*. Then, the approximate volume of the shell is

$$V_{shell} \approx 2\pi(x_i^* + k)f(x_i^*)\Delta x.$$

The remainder of the development proceeds as before, and we see that

$$V = \int_a^b (2\pi(x + k)f(x))dx.$$

We could also rotate the region around other horizontal or vertical lines, such as a vertical line in the right half plane. In each case, the volume formula must be adjusted accordingly. Specifically, the x-term in the integral must be replaced with an expression representing the radius of a shell. To see how this works, consider the following example.

Example 2.3.4: A Region of Revolution Revolved around a Line

Define R as the region bounded above by the graph of $f(x) = x$ and below by the x-axis over the interval $[1, 2]$. Find the volume of the solid of revolution formed by revolving R around the line $x = -1$.

Solution

First, graph the region R and the associated solid of revolution, as shown in Figure 2.3.8.

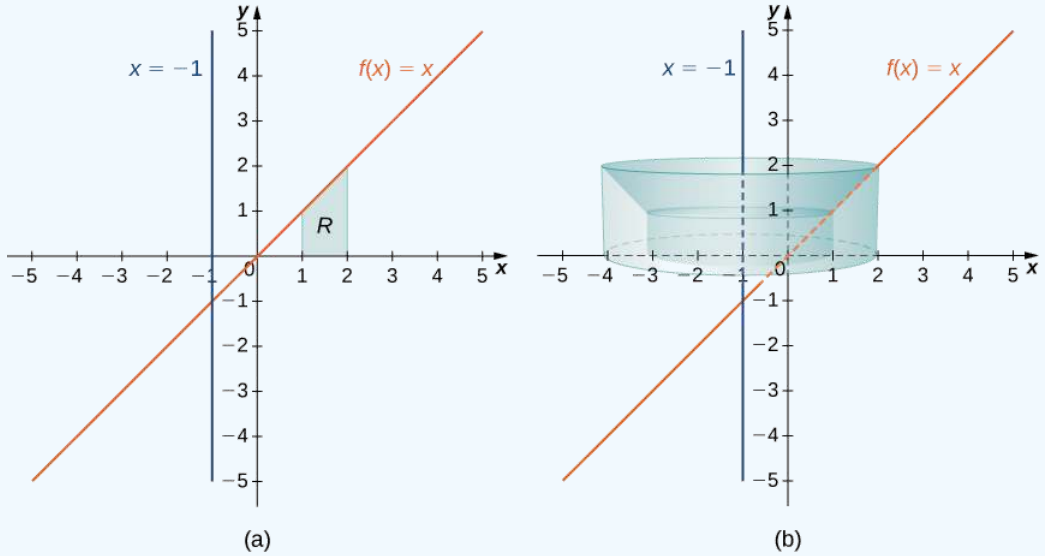


Figure 2.3.8: (a) The region R between the graph of $f(x)$ and the x-axis over the interval $[1,2]$. (b) The solid of revolution generated by revolving R around the line $x=-1$.

Note that the radius of a shell is given by $x+1$. Then the volume of the solid is given by

$$\begin{aligned} V &= \int_1^2 2\pi(x+1)f(x)\,dx \\ &= \int_1^2 2\pi(x+1)x\,dx = 2\pi\int_1^2 x^2+x\,dx \\ &= 2\pi\left[\frac{x^3}{3}+\frac{x^2}{2}\right]\Big|_1^2 \\ &= \frac{23\pi}{3}\text{ units}^3 \end{aligned}$$

? Exercise 2.3.4

Define R as the region bounded above by the graph of $f(x)=x^2$ and below by the x-axis over the interval $[0,1]$. Find the volume of the solid of revolution formed by revolving R around the line $x=-2$.

Hint

Use the process from Example 2.3.4

Answer

$\frac{11\pi}{6}$ units^3

For our final example in this section, let's look at the volume of a solid of revolution for which the region of revolution is bounded by the graphs of two functions.

✔ Example 2.3.5: A Region of Revolution Bounded by the Graphs of Two Functions

Define R as the region bounded above by the graph of the function $f(x)=\sqrt{x}$ and below by the graph of the function $g(x)=1/x$ over the interval $[1,4]$. Find the volume of the solid of revolution generated by revolving R around the y-axis.

Solution

First, graph the region R and the associated solid of revolution, as shown in Figure 2.3.9.

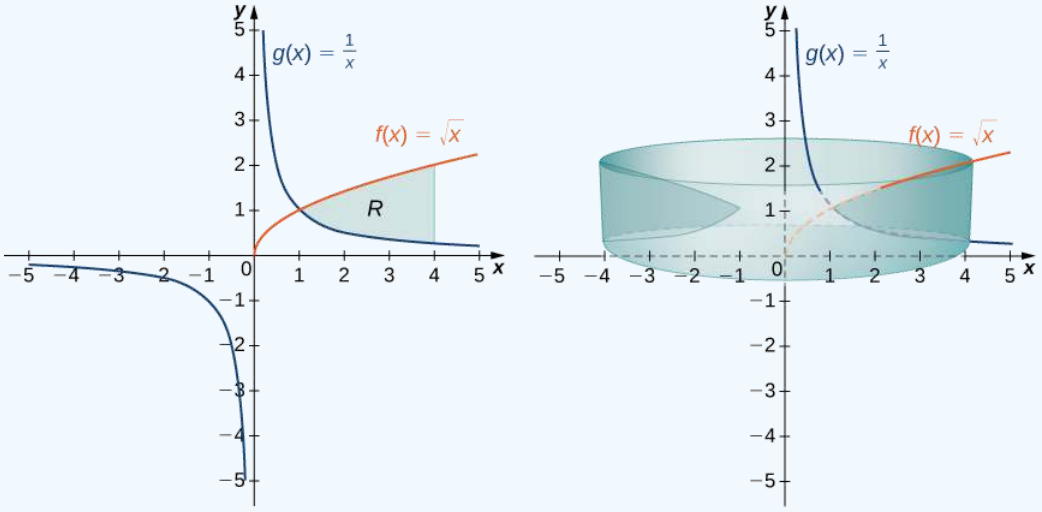


Figure 2.3.9: (a) The region R between the graph of $f(x)$ and the graph of $g(x)$ over the interval $[1,4]$. (b) The solid of revolution generated by revolving R around the y-axis.

Note that the axis of revolution is the y-axis, so the radius of a shell is given simply by x. We don't need to make any adjustments to the x-term of our integrand. The height of a shell, though, is given by $f(x)-g(x)$, so in this case we need to adjust the $f(x)$ term of the integrand. Then the volume of the solid is given by

$$V = \int_1^4 (2\pi\, x(f(x)-g(x)))\, dx$$

$$= \int_1^4 (2\pi\, x(\sqrt{x} - \frac{1}{x}))\, dx = 2\pi \int_1^4 (x^{3/2} - 1) dx$$

$$= 2\pi \left[\frac{2x^{5/2}}{5} - x\right]\Big|_1^4 = \frac{94\pi}{5} \text{ units}^3.$$

? Exercise 2.3.5

Define R as the region bounded above by the graph of $f(x) = x$ and below by the graph of $g(x) = x^2$ over the interval $[0,1]$. Find the volume of the solid of revolution formed by revolving R around the y-axis.

Hint

Hint: Use the process from Example 2.3.5

Answer

$\frac{\pi}{6}$ units^3

Which Method Should We Use?

We have studied several methods for finding the volume of a solid of revolution, but how do we know which method to use? It often comes down to a choice of which integral is easiest to evaluate. Figure 2.3.10 describes the different approaches for solids of revolution around the x-axis. It's up to you to develop the analogous table for solids of revolution around the y-axis.

Comparing the Methods for Finding the Volume of a Solid Revolution around the x-axis

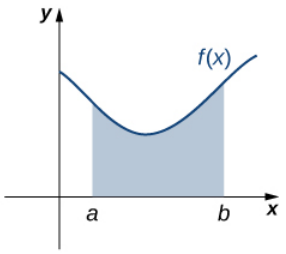
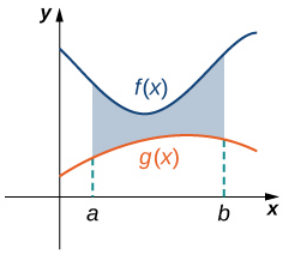
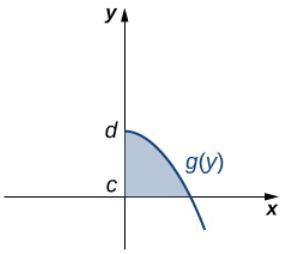
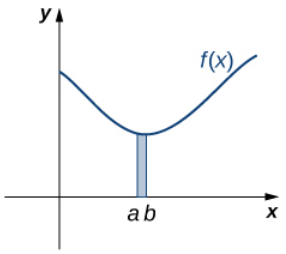
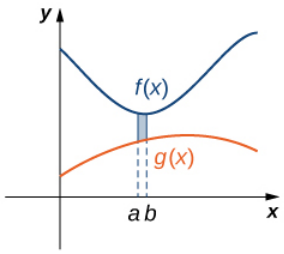
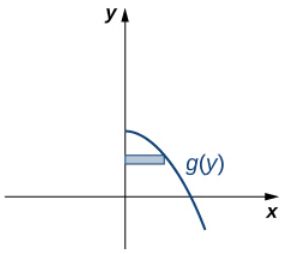
Compare	Disk Method	Washer Method	Shell Method
Volume formula	$V = \int_a^b \pi[f(x)]^2\,dx$	$V = \int_a^b \pi\left[(f(x))^2 - (g(x))^2\right]dx$	$V = \int_c^d 2\pi y\, g(y)\,dy$
Solid	No cavity in the center	Cavity in the center	With or without a cavity in the center
Interval to partition	$[a, b]$ on x-axis	$[a, b]$ on x-axis	$[c, d]$ on y-axis
Rectangle	Vertical	Vertical	Horizontal
Typical region			
Typical element			

Figure 2.3.10

Let's take a look at a couple of additional problems and decide on the best approach to take for solving them.

Example 2.3.6: Selecting the Best Method

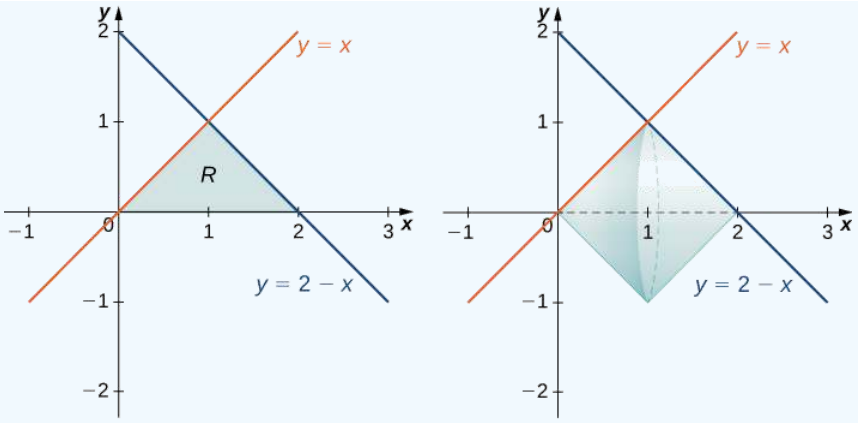
For each of the following problems, select the best method to find the volume of a solid of revolution generated by revolving the given region around the x-axis, and set up the integral to find the volume (do not evaluate the integral).

a. The region bounded by the graphs of $y = x, y = 2 - x,$ and the x-axis.
b. The region bounded by the graphs of $y = 4x - x^2$ and the x-axis.

Solution

a.

First, sketch the region and the solid of revolution as shown.



(a) (b)

Figure 2.3.11: (a) The region R bounded by two lines and the x-axis. (b) The solid of revolution generated by revolving R about the x-axis.

Looking at the region, if we want to integrate with respect to x, we would have to break the integral into two pieces, because we have different functions bounding the region over $[0, 1]$ and $[1, 2]$. In this case, using the disk method, we would have

$$V = \int_0^1 \pi x^2 \, dx + \int_1^2 \pi (2 - x)^2 \, dx.$$

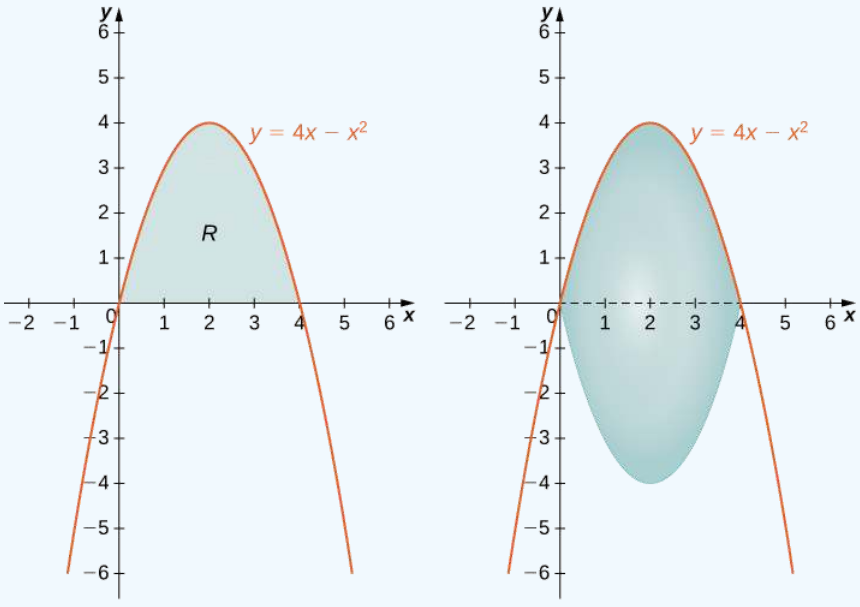
If we used the shell method instead, we would use functions of y to represent the curves, producing

$$V = \int_0^1 2\pi y[(2 - y) - y] \, dy = \int_0^1 2\pi y[2 - 2y] \, dy.$$

Neither of these integrals is particularly onerous, but since the shell method requires only one integral, and the integrand requires less simplification, we should probably go with the shell method in this case.

b.

First, sketch the region and the solid of revolution as shown.



(a) (b)

Figure 2.3.12: (a) The region R between the curve and the x-axis. (b) The solid of revolution generated by revolving R about the x-axis.

Looking at the region, it would be problematic to define a horizontal rectangle; the region is bounded on the left and right by the same function. Therefore, we can dismiss the method of shells. The solid has no cavity in the middle, so we can use the method of disks. Then

$$V = \int_0^4 \pi \left(4x - x^2\right)^2 \, dx$$

? Exercise 2.3.6

Select the best method to find the volume of a solid of revolution generated by revolving the given region around the x-axis, and set up the integral to find the volume (do not evaluate the integral): the region bounded by the graphs of $y = 2 - x^2$ and $y = x^2$.

Hint

Sketch the region and use Figure 2.3.12to decide which integral is easiest to evaluate.

Answer

Use the method of washers;

$$V = \int_{-1}^1 \pi \left[\left(2 - x^2\right)^2 - \left(x^2\right)^2 \right] \, dx$$

Key Concepts

- The method of cylindrical shells is another method for using a definite integral to calculate the volume of a solid of revolution. This method is sometimes preferable to either the method of disks or the method of washers because we integrate with respect to the other variable. In some cases, one integral is substantially more complicated than the other.
- The geometry of the functions and the difficulty of the integration are the main factors in deciding which integration method to use.

Key Equations

- **Method of Cylindrical Shells**

$$V = \int_a^b \left(2\pi \, x \, f(x)\right) \, dx$$

Glossary

method of cylindrical shells

a method of calculating the volume of a solid of revolution by dividing the solid into nested cylindrical shells; this method is different from the methods of disks or washers in that we integrate with respect to the opposite variable

This page titled 2.3: Volumes of Revolution - Cylindrical Shells is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.3: Volumes of Revolution - Cylindrical Shells** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.4: Arc Length of a Curve and Surface Area

Learning Objectives

- Determine the length of a curve, $y = f(x)$, between two points.
- Determine the length of a curve, $x = g(y)$, between two points.
- Find the surface area of a solid of revolution.

In this section, we use definite integrals to find the arc length of a curve. We can think of **arc length** as the distance you would travel if you were walking along the path of the curve. Many real-world applications involve arc length. If a rocket is launched along a parabolic path, we might want to know how far the rocket travels. Or, if a curve on a map represents a road, we might want to know how far we have to drive to reach our destination.

We begin by calculating the arc length of curves defined as functions of x, then we examine the same process for curves defined as functions of y. (The process is identical, with the roles of x and y reversed.) The techniques we use to find arc length can be extended to find the surface area of a surface of revolution, and we close the section with an examination of this concept.

Arc Length of the Curve y = f(x)

In previous applications of integration, we required the function $f(x)$ to be integrable, or at most continuous. However, for calculating arc length we have a more stringent requirement for $f(x)$. Here, we require $f(x)$ to be differentiable, and furthermore we require its derivative, $f'(x)$, to be continuous. Functions like this, which have continuous derivatives, are called **smooth.** (This property comes up again in later chapters.)

Let $f(x)$ be a smooth function defined over $[a, b]$. We want to calculate the length of the curve from the point $(a, f(a))$ to the point $(b, f(b))$. We start by using line segments to approximate the length of the curve. For $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of $[a, b]$. Then, for $i = 1, 2, \ldots, n$, construct a line segment from the point $(x_{i-1}, f(x_{i-1}))$ to the point $(x_i, f(x_i))$. Although it might seem logical to use either horizontal or vertical line segments, we want our line segments to approximate the curve as closely as possible. Figure 2.4.1 depicts this construct for $n = 5$.

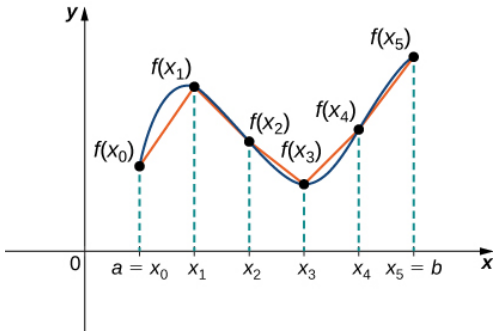


Figure 2.4.1: We can approximate the length of a curve by adding line segments.

To help us find the length of each line segment, we look at the change in vertical distance as well as the change in horizontal distance over each interval. Because we have used a regular partition, the change in horizontal distance over each interval is given by Δx. The change in vertical distance varies from interval to interval, though, so we use $\Delta y_i = f(x_i) - f(x_{i-1})$ to represent the change in vertical distance over the interval $[x_{i-1}, x_i]$, as shown in Figure 2.4.2. Note that some (or all) Δy_i may be negative.

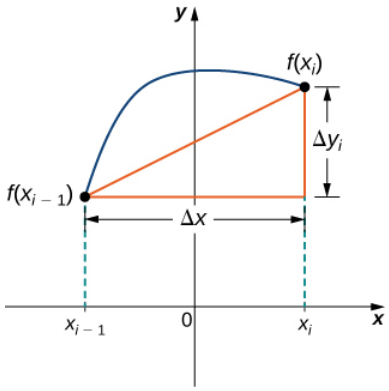


Figure 2.4.2: A representative line segment approximates the curve over the interval $[x_{i-1}, x_i]$.

By the Pythagorean theorem, the length of the line segment is

$$\sqrt{(\Delta x)^2 + (\Delta y_i)^2}.$$

We can also write this as

$$\Delta x \sqrt{1 + ((\Delta y_i)/(\Delta x))^2}.$$

Now, by the Mean Value Theorem, there is a point $x_i^* \in [x_{i-1}, x_i]$ such that $f'(x_i^*) = (\Delta y_i)/(\Delta x)$. Then the length of the line segment is given by

$$\Delta x \sqrt{1 + [f'(x_i^*)]^2}.$$

Adding up the lengths of all the line segments, we get

$$\text{Arc Length} \approx \sum_{i=1}^{n} \sqrt{1 + [f'(x_i^*)]^2}\,\Delta x.$$

This is a Riemann sum. Taking the limit as $n \to \infty$, we have

$$\begin{aligned} \text{Arc Length} &= \lim_{n\to\infty} \sum_{i=1}^{n} \sqrt{1 + [f'(x_i^*)]^2}\,\Delta x \\ &= \int_a^b \sqrt{1 + [f'(x)]^2}\,dx. \end{aligned}$$

We summarize these findings in the following theorem.

Arc Length for $y = f(x)$

Let $f(x)$ be a smooth function over the interval $[a, b]$. Then the arc length of the portion of the graph of $f(x)$ from the point $(a, f(a))$ to the point $(b, f(b))$ is given by

$$\text{Arc Length} = \int_a^b \sqrt{1 + [f'(x)]^2}\,dx.$$

Note that we are integrating an expression involving $f'(x)$, so we need to be sure $f'(x)$ is integrable. This is why we require $f(x)$ to be smooth. The following example shows how to apply the theorem.

Example 2.4.1: Calculating the Arc Length of a Function of x

Let $f(x) = 2x^{3/2}$. Calculate the arc length of the graph of $f(x)$ over the interval $[0, 1]$. Round the answer to three decimal places.

Solution

We have $f'(x) = 3x^{1/2}$, so $[f'(x)]^2 = 9x$. Then, the arc length is

$$\text{Arc Length} = \int_a^b \sqrt{1+[f'(x)]^2}dx$$
$$= \int_0^1 \sqrt{1+9x}\, dx.$$

Substitute $u = 1+9x$. Then, $du = 9\, dx$. When $x = 0$, then $u = 1$, and when $x = 1$, then $u = 10$. Thus,

$$\text{Arc Length} = \int_0^1 \sqrt{1+9x}\, dx$$
$$= \frac{1}{9}\int_0^1 \sqrt{1+9x}\cdot 9\, dx$$
$$= \frac{1}{9}\int_1^{10} \sqrt{u}\, du$$
$$= \frac{1}{9}\cdot\frac{2}{3}u^{3/2}\Big|_1^{10} = \frac{2}{27}[10\sqrt{10}-1]$$
$$\approx 2.268 \text{ units}.$$

Exercise 2.4.1

Let $f(x) = (4/3)x^{3/2}$. Calculate the arc length of the graph of $f(x)$ over the interval $[0, 1]$. Round the answer to three decimal places.

Hint

Use the process from the previous example. Don't forget to change the limits of integration.

Answer

$$\frac{1}{6}(5\sqrt{5}-1) \approx 1.697 \text{ units}$$

Although it is nice to have a formula for calculating arc length, this particular theorem can generate expressions that are difficult to integrate. We study some techniques for integration in Introduction to Techniques of Integration. In some cases, we may have to use a computer or calculator to approximate the value of the integral.

Example 2.4.2: Using a Computer or Calculator to Determine the Arc Length of a Function of x

Let $f(x) = x^2$. Calculate the arc length of the graph of $f(x)$ over the interval $[1, 3]$.

Solution

We have $f'(x) = 2x$, so $[f'(x)]^2 = 4x^2$. Then the arc length is given by

$$\text{Arc Length} = \int_a^b \sqrt{1+[f'(x)]^2}\, dx$$
$$= \int_1^3 \sqrt{1+4x^2}\, dx.$$

Using a computer to approximate the value of this integral, we get

$$\int_1^3 \sqrt{1+4x^2}\, dx \approx 8.26815 \text{ units}.$$

Exercise 2.4.2

Let $f(x) = \sin x$. Calculate the arc length of the graph of $f(x)$ over the interval $[0, \pi]$. Use a computer or calculator to approximate the value of the integral.

Hint

Use the process from the previous example.

Answer

$$\text{Arc Length} \approx 3.8202 \text{ units}$$

Arc Length of the Curve x = g(y)

We have just seen how to approximate the length of a curve with line segments. If we want to find the arc length of the graph of a function of y, we can repeat the same process, except we partition the y-axis instead of the x-axis. Figure 2.4.3 shows a representative line segment.

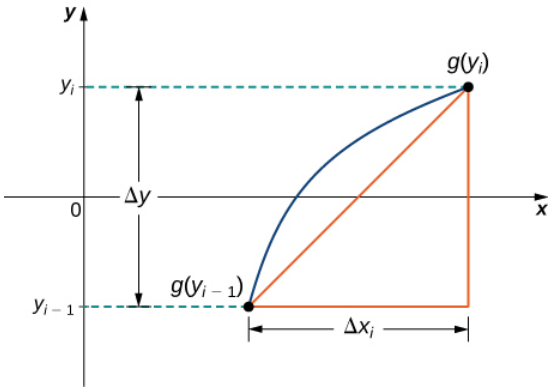


Figure 2.4.3: A representative line segment over the interval $[y_{i-1}, y_i]$.

Then the length of the line segment is

$$\sqrt{(\Delta y)^2 + (\Delta x_i)^2},$$

which can also be written as

$$\Delta y \sqrt{1 + \left(\frac{\Delta x_i}{\Delta y}\right)^2}.$$

If we now follow the same development we did earlier, we get a formula for arc length of a function $x = g(y)$.

Arc Length for $x = g(y)$

Let $g(y)$ be a smooth function over an interval $[c, d]$. Then, the arc length of the graph of $g(y)$ from the point $(c, g(c))$ to the point $(d, g(d))$ is given by

$$\text{Arc Length} = \int_c^d \sqrt{1 + [g'(y)]^2}\, dy.$$

Example 2.4.3: Calculating the Arc Length of a Function of y

Let $g(y) = 3y^3$. Calculate the arc length of the graph of $g(y)$ over the interval $[1, 2]$.

Solution

We have $g'(y) = 9y^2$, so $[g'(y)]^2 = 81y^4$. Then the arc length is

$$\text{Arc Length} = \int_c^d \sqrt{1+[g'(y)]^2}\, dy$$
$$= \int_1^2 \sqrt{1+81y^4}\, dy.$$

Using a computer to approximate the value of this integral, we obtain

$$\int_1^2 \sqrt{1+81y^4}dy \approx 21.0277 \text{ units}.$$

? Exercise 2.4.3

Let $g(y) = 1/y$. Calculate the arc length of the graph of $g(y)$ over the interval $[1, 4]$. Use a computer or calculator to approximate the value of the integral.

Hint

Use the process from the previous example.

Answer

$$\text{Arc Length} = 3.15018 \text{ units}$$

Area of a Surface of Revolution

The concepts we used to find the arc length of a curve can be extended to find the surface area of a surface of revolution. **Surface area** is the total area of the outer layer of an object. For objects such as cubes or bricks, the surface area of the object is the sum of the areas of all of its faces. For curved surfaces, the situation is a little more complex. Let $f(x)$ be a nonnegative smooth function over the interval $[a, b]$. We wish to find the surface area of the surface of revolution created by revolving the graph of $y = f(x)$ around the x-axis as shown in the following figure.

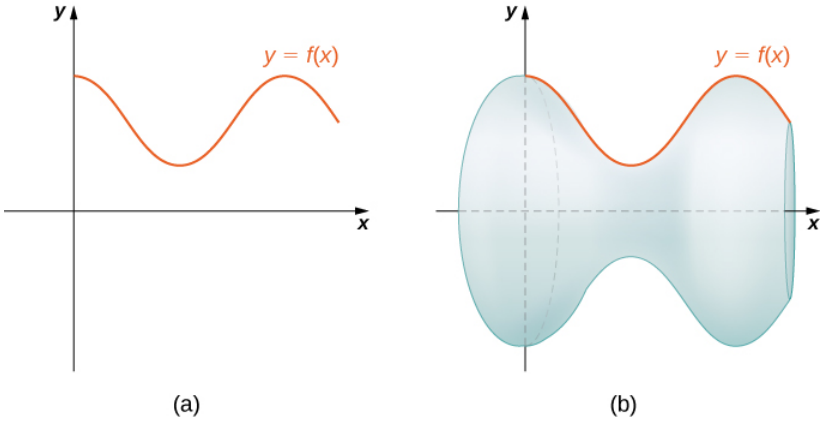


Figure 2.4.4: (a) A curve representing the function $f(x)$. (b) The surface of revolution formed by revolving the graph of $f(x)$ around the $x-axis$.

As we have done many times before, we are going to partition the interval $[a, b]$ and approximate the surface area by calculating the surface area of simpler shapes. We start by using line segments to approximate the curve, as we did earlier in this section. For $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of $[a, b]$. Then, for $i = 1, 2, \ldots, n$, construct a line segment from the point $(x_{i-1}, f(x_{i-1}))$ to the point $(x_i, f(x_i))$. Now, revolve these line segments around the x-axis to generate an approximation of the surface of revolution as shown in the following figure.

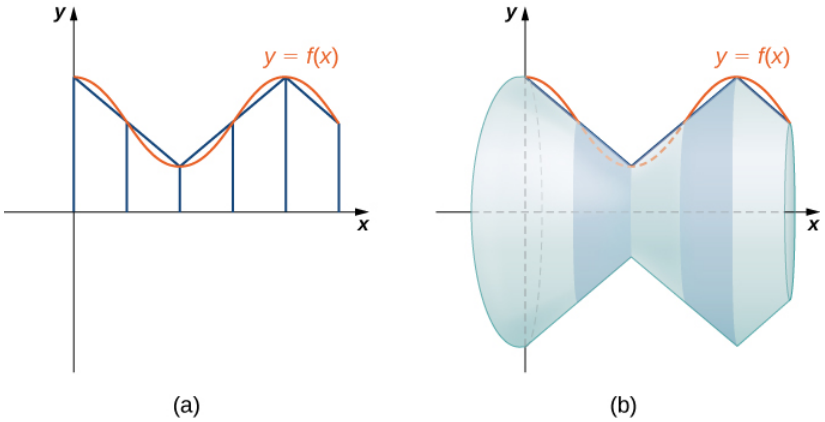


Figure 2.4.5: (a) Approximating $f(x)$ with line segments. (b) The surface of revolution formed by revolving the line segments around the x-axis.

Notice that when each line segment is revolved around the axis, it produces a band. These bands are actually pieces of cones (think of an ice cream cone with the pointy end cut off). A piece of a cone like this is called a **frustum** of a cone.

To find the surface area of the band, we need to find the lateral surface area, S, of the frustum (the area of just the slanted outside surface of the frustum, not including the areas of the top or bottom faces). Let r_1 and r_2 be the radii of the wide end and the narrow end of the frustum, respectively, and let l be the slant height of the frustum as shown in the following figure.

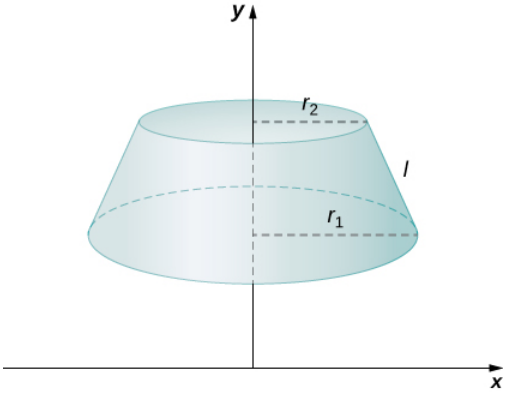


Figure 2.4.6: A frustum of a cone can approximate a small part of surface area.

We know the lateral surface area of a cone is given by

$$\text{Lateral Surface Area } = \pi r s,$$

where r is the radius of the base of the cone and s is the **slant height** (Figure 2.4.7).

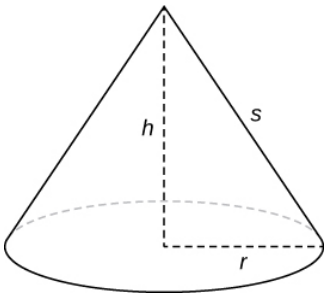


Figure 2.4.7: The lateral surface area of the cone is given by $\pi r s$.

Since a frustum can be thought of as a piece of a cone, the lateral surface area of the frustum is given by the lateral surface area of the whole cone less the lateral surface area of the smaller cone (the pointy tip) that was cut off (Figure 2.4.8).

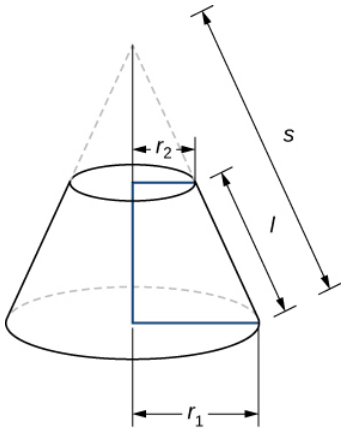


Figure 2.4.8: Calculating the lateral surface area of a frustum of a cone.

The cross-sections of the small cone and the large cone are similar triangles, so we see that

$$\frac{r_2}{r_1} = \frac{s-l}{s}$$

Solving for s, we get

$$\begin{aligned} \frac{r_2}{r_1} &= \frac{s-l}{s} \\ r_2 s &= r_1(s-l) \\ r_2 s &= r_1 s - r_1 l \\ r_1 l &= r_1 s - r_2 s \\ r_1 l &= (r_1 - r_2)s \\ \frac{r_1 l}{r_1 - r_2} &= s \end{aligned}$$

Then the lateral surface area (SA) of the frustum is

$$\begin{aligned} S &= (\text{Lateral SA of large cone}) - (\text{Lateral SA of small cone}) \\ &= \pi r_1 s - \pi r_2 (s-l) \\ &= \pi r_1 \left(\frac{r_1 l}{r_1 - r_2}\right) - \pi r_2 \left(\frac{r_1 l}{r_1 - r_2} - l\right) \\ &= \frac{\pi r_1^2 l}{r^1 - r^2} - \frac{\pi r_1 r_2 l}{r_1 - r_2} + \pi r_2 l \\ &= \frac{\pi r_1^2 l}{r_1 - r_2} - \frac{\pi r_1 r 2_l}{r_1 - r_2} + \frac{\pi r_2 l (r_1 - r_2)}{r_1 - r_2} \\ &= \frac{\pi r_1^2}{l r_1 - r_2} - \frac{\pi r_1 r_2 l}{r_1 - r_2} + \frac{\pi r_1 r_2 l}{r_1 - r_2} - \frac{\pi r_2^2 l}{r_1 - r_3} \\ &= \frac{\pi(r_1^2 - r_2^2)l}{r_1 - r_2} = \frac{\pi(r_1 - r + 2)(r1 + r2)l}{r_1 - r_2} \\ &= \pi(r_1 + r_2)l. \end{aligned}$$

Let's now use this formula to calculate the surface area of each of the bands formed by revolving the line segments around the x-axis. A representative band is shown in the following figure.

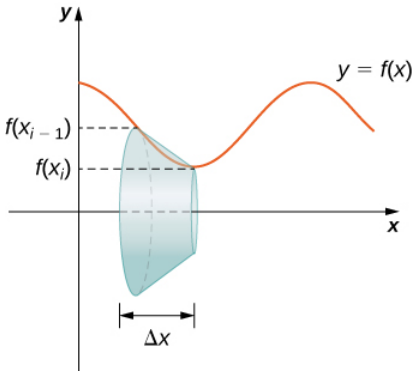


Figure 2.4.9: A representative band used for determining surface area.

Note that the slant height of this frustum is just the length of the line segment used to generate it. So, applying the surface area formula, we have

$$\begin{aligned} S &= \pi(r_1 + r_2)l \\ &= \pi\big(f(x_{i-1}) + f(x_i)\big)\sqrt{\Delta x^2 + (\Delta y_i)^2} \\ &= \pi\big(f(x_{i-1}) + f(x_i)\big)\Delta x\sqrt{1 + \left(\frac{\Delta y_i}{\Delta x}\right)^2} \end{aligned}$$

Now, as we did in the development of the arc length formula, we apply the Mean Value Theorem to select $x_i^* \in [x_{i-1}, x_i]$ such that $f\,'(x_i^*) = (\Delta y_i)/\Delta x$. This gives us

$$S = \pi\big(f(x_{i-1}) + f(x_i)\big)\Delta x\sqrt{1 + \big(f\,'(x_i^*)\big)^2}$$

Furthermore, since $f(x)$ is continuous, by the *Intermediate Value Theorem*, there is a point $x_i^{**} \in [x_{i-1}, x_i]$ such that $f(x_i^{**}) = \frac{1}{2}\left[f(x_{i-1}) + f(x_i)\right]$,

so we get

$$S = 2\pi f(x_i^{**})\Delta x\sqrt{1 + \big(f\,'(x_i^*)\big)^2}.$$

Then the approximate surface area of the whole surface of revolution is given by

$$\text{Surface Area} \approx \sum_{i=1}^{n} 2\pi f(x_i^{**})\Delta x\sqrt{1 + \big(f\,'(x_i^*)\big)^2}.$$

This almost looks like a Riemann sum, except we have functions evaluated at two different points, x_i^* and x_i^{**}, over the interval $[x_{i-1}, x_i]$. Although we do not examine the details here, it turns out that because $f(x)$ is smooth, if we let $n \to \infty$, the limit works the same as a Riemann sum even with the two different evaluation points. This makes sense intuitively. Both x_i^* and x_i^{**} are in the interval $[x_{i-1}, x_i]$, so it makes sense that as $n \to \infty$, both x_i^* and x_i^{**} approach x Those of you who are interested in the details should consult an advanced calculus text.

Taking the limit as $n \to \infty$, we get

$$\begin{aligned} \text{Surface Area} &= \lim_{n\to\infty} \sum_{i=1}^{n} 2\pi f(x_i^{**})\Delta x\sqrt{1 + \big(f\,'(x_i^*)\big)^2} \\ &= \int_a^b \left(2\pi f(x)\sqrt{1 + \big(f\,'(x)\big)^2}\right) dx \end{aligned}$$

As with arc length, we can conduct a similar development for functions of y to get a formula for the surface area of surfaces of revolution about the y-axis. These findings are summarized in the following theorem.

Surface Area of a Surface of Revolution

Let $f(x)$ be a nonnegative smooth function over the interval $[a, b]$. Then, the surface area of the surface of revolution formed by revolving the graph of $f(x)$ around the x-axis is given by

$$\text{Surface Area} = \int_a^b \left(2\pi f(x)\sqrt{1+\left(f'(x)\right)^2}\right) dx$$

Similarly, let $g(y)$ be a nonnegative smooth function over the interval $[c, d]$. Then, the surface area of the surface of revolution formed by revolving the graph of $g(y)$ around the y-axis is given by

$$\text{Surface Area} = \int_c^d \left(2\pi g(y)\sqrt{1+\left(g'(y)\right)^2}\right) dy$$

Example 2.4.4: Calculating the Surface Area of a Surface of Revolution 1.

Let $f(x) = \sqrt{x}$ over the interval $[1, 4]$. Find the surface area of the surface generated by revolving the graph of $f(x)$ around the x-axis. Round the answer to three decimal places.

Solution

The graph of $f(x)$ and the surface of rotation are shown in Figure 2.4.10.

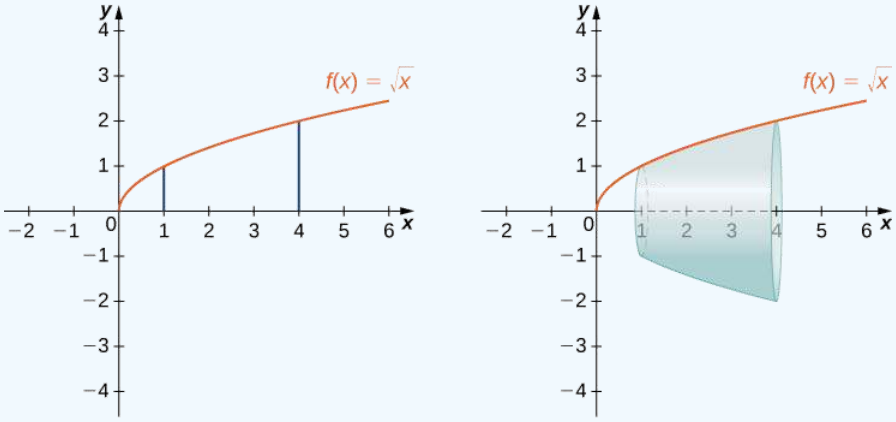


Figure 2.4.10: (a) The graph of $f(x)$. (b) The surface of revolution.

We have $f(x) = \sqrt{x}$. Then, $f'(x) = \frac{1}{2\sqrt{x}}$ and $\left(f'(x)\right)^2 = \frac{1}{4x}$. Then,

$$\begin{aligned}\text{Surface Area} &= \int_a^b \left(2\pi f(x)\sqrt{1+\left(f'(x)\right)^2}\right) dx \\ &= \int_1^4 \left(2\pi\sqrt{x}\sqrt{1+\frac{1}{4x}}\right) dx \\ &= \int_1^4 \left(2\pi\sqrt{x+\frac{1}{4}}\right) dx.\end{aligned}$$

Let $u = x + 1/4$. Then, $du = dx$. When $x = 1,\ u = 5/4$, and when $x = 4,\ u = 17/4$. This gives us

$$\begin{aligned}\int_0^1 \left(2\pi\sqrt{x+\frac{1}{4}}\right) dx &= \int_{5/4}^{17/4} 2\pi\sqrt{u}\, du \\ &= 2\pi\left[\frac{2}{3}u^{3/2}\right]\Big|_{5/4}^{17/4} \\ &= \frac{\pi}{6}[17\sqrt{17} - 5\sqrt{5}] \approx 30.846 \text{ units}^2\end{aligned}$$

Exercise 2.4.4

Let $f(x) = \sqrt{1-x}$ over the interval $[0, 1/2]$. Find the surface area of the surface generated by revolving the graph of $f(x)$ around the x-axis. Round the answer to three decimal places.

Hint

Use the process from the previous example.

Answer

$$\frac{\pi}{6}(5\sqrt{5} - 3\sqrt{3}) \approx 3.133 \text{ units}^2$$

Example 2.4.5: Calculating the Surface Area of a Surface of Revolution 2

Let $f(x) = y = \sqrt[3]{3x}$. Consider the portion of the curve where $0 \le y \le 2$. Find the surface area of the surface generated by revolving the graph of $f(x)$ around the y-axis.

Solution

Notice that we are revolving the curve around the y-axis, and the interval is in terms of y, so we want to rewrite the function as a function of y. We get $x = g(y) = (1/3)y^3$. The graph of $g(y)$ and the surface of rotation are shown in the following figure.

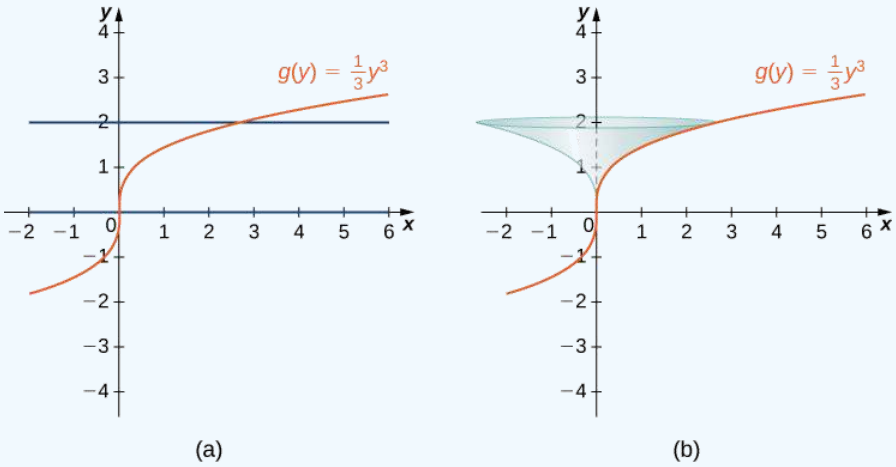


Figure 2.4.11: (a) The graph of $g(y)$. (b) The surface of revolution.

We have $g(y) = (1/3)y^3$, so $g'(y) = y^2$ and $\left(g'(y)\right)^2 = y^4$. Then

$$\begin{aligned} \text{Surface Area} &= \int_c^d \left(2\pi g(y)\sqrt{1+\left(g'(y)\right)^2}\right) dy \\ &= \int_0^2 \left(2\pi\left(\frac{1}{3}y^3\right)\sqrt{1+y^4}\right) dy \\ &= \frac{2\pi}{3}\int_0^2 \left(y^3\sqrt{1+y^4}\right) dy. \end{aligned}$$

Let $u = y^4 + 1$. Then $du = 4y^3\, dy$. When $y = 0$, $u = 1$, and when $y = 2$, $u = 17$. Then

$$\begin{aligned} \frac{2\pi}{3}\int_0^2 \left(y^3\sqrt{1+y^4}\right) dy &= \frac{2\pi}{3}\int_1^{17} \frac{1}{4}\sqrt{u}\, du \\ &= \frac{\pi}{6}\left[\frac{2}{3}u^{3/2}\right]\Big|_1^{17} = \frac{\pi}{9}\left[(17)^{3/2} - 1\right] \approx 24.118 \text{ units}^2. \end{aligned}$$

? Exercise 2.4.5

Let $g(y) = \sqrt{9 - y^2}$ over the interval $y \in [0, 2]$. Find the surface area of the surface generated by revolving the graph of $g(y)$ around the y-axis.

Hint

Use the process from the previous example.

Answer

12π units^2

Key Concepts

- The arc length of a curve can be calculated using a definite integral.
- The arc length is first approximated using line segments, which generates a Riemann sum. Taking a limit then gives us the definite integral formula. The same process can be applied to functions of y.
- The concepts used to calculate the arc length can be generalized to find the surface area of a surface of revolution.
- The integrals generated by both the arc length and surface area formulas are often difficult to evaluate. It may be necessary to use a computer or calculator to approximate the values of the integrals.

Key Equations

- **Arc Length of a Function of x**

$$\textbf{Arc Length} = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx$$

- **Arc Length of a Function of y**

$$\textbf{Arc Length} = \int_c^d \sqrt{1 + [g'(y)]^2}\, dy$$

- **Surface Area of a Function of x**

$$\textbf{Surface Area} = \int_a^b \left(2\pi f(x)\sqrt{1 + \left(f'(x)\right)^2}\right) dx$$

Glossary

arc length

the arc length of a curve can be thought of as the distance a person would travel along the path of the curve

frustum

a portion of a cone; a frustum is constructed by cutting the cone with a plane parallel to the base

surface area

the surface area of a solid is the total area of the outer layer of the object; for objects such as cubes or bricks, the surface area of the object is the sum of the areas of all of its faces

---

This page titled 2.4: Arc Length of a Curve and Surface Area is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.4: Arc Length of a Curve and Surface Area** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.5: Physical Applications of Integration

Learning Objectives

- Determine the mass of a one-dimensional object from its linear density function.
- Determine the mass of a two-dimensional circular object from its radial density function.
- Calculate the work done by a variable force acting along a line.
- Calculate the work done in pumping a liquid from one height to another.
- Find the hydrostatic force against a submerged vertical plate.

In this section, we examine some physical applications of integration. Let's begin with a look at calculating mass from a density function. We then turn our attention to work, and close the section with a study of hydrostatic force.

Mass and Density

We can use integration to develop a formula for calculating mass based on a density function. First we consider a thin rod or wire. Orient the rod so it aligns with the x-axis, with the left end of the rod at $x = a$ and the right end of the rod at $x = b$ (Figure 2.5.1). Note that although we depict the rod with some thickness in the figures, for mathematical purposes we assume the rod is thin enough to be treated as a one-dimensional object.

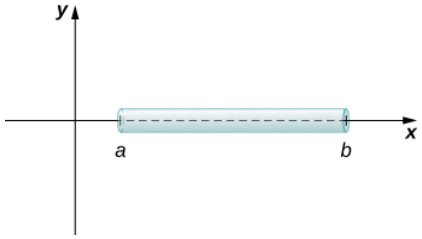


Figure 2.5.1: We can calculate the mass of a thin rod oriented along the x-axis by integrating its density function.

If the rod has constant density ρ, given in terms of mass per unit length, then the mass of the rod is just the product of the density and the length of the rod: $(b-a)\rho$. If the density of the rod is not constant, however, the problem becomes a little more challenging. When the density of the rod varies from point to point, we use a linear density function, $\rho(x)$, to denote the density of the rod at any point, x. Let $\rho(x)$ be an integrable linear density function. Now, for $i = 0, 1, 2, \ldots, n$ let $P = x_i$ be a regular partition of the interval $[a, b]$, and for $i = 1, 2, \ldots, n$ choose an arbitrary point $x_i^* \in [x_{i-1}, x_i]$. Figure 2.5.2 shows a representative segment of the rod.

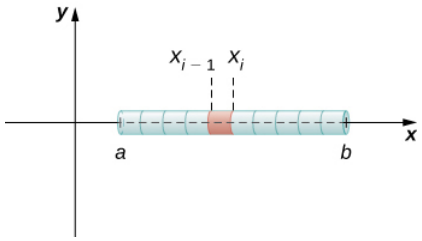


Figure 2.5.2: A representative segment of the rod.

The mass m_i of the segment of the rod from x_{i-1} to x_i is approximated by

$$\begin{aligned} m_i &\approx \rho(x_i^*)(x_i - x_{i-1}) \\ &= \rho(x_i^*)\Delta x. \end{aligned}$$

Adding the masses of all the segments gives us an approximation for the mass of the entire rod:

$$\begin{aligned} m &= \sum_{i=1}^{n} m_i \\ &\approx \sum_{i=1}^{n} \rho(x_i^*)\Delta x. \end{aligned}$$

This is a Riemann sum. Taking the limit as $n \to \infty$, we get an expression for the exact mass of the rod:

$$\begin{aligned} m &= \lim_{n\to\infty} \sum_{i=1}^{n} \rho(x_i^*)\Delta x \\ &= \int_a^b \rho(x)dx. \end{aligned}$$

We state this result in the following theorem.

Mass–Density Formula of a One-Dimensional Object

Given a thin rod oriented along the x-axis over the interval $[a, b]$, let $\rho(x)$ denote a linear density function giving the density of the rod at a point x in the interval. Then the mass of the rod is given by

$$m = \int_a^b \rho(x)dx. \tag{2.5.1}$$

We apply this theorem in the next example.

Example 2.5.1: Calculating Mass from Linear Density

Consider a thin rod oriented on the x-axis over the interval $[\pi/2, \pi]$ If the density of the rod is given by $\rho(x) = \sin x$, what is the mass of the rod?

Solution

Applying Equation 2.5.1 directly, we have

$$\begin{aligned} m &= \int_a^b \rho(x)dx \\ &= \int_{\pi/2}^{\pi} \sin x\, dx \\ &= -\cos x\Big|_{\pi/2}^{\pi} \\ &= 1. \end{aligned}$$

Exercise 2.5.1

Consider a thin rod oriented on the x-axis over the interval $[1, 3]$. If the density of the rod is given by $\rho(x) = 2x^2 + 3$, what is the mass of the rod?

Hint

Use the process from the previous example.

Solution

$70/3$

We now extend this concept to find the mass of a two-dimensional disk of radius r. As with the rod we looked at in the one-dimensional case, here we assume the disk is thin enough that, for mathematical purposes, we can treat it as a two-dimensional object. We assume the density is given in terms of mass per unit area (called **area density**), and further assume the density varies only along the disk's radius (called **radial density**). We orient the disk in the $xy-plane$, with the center at the origin. Then, the density of the disk can be treated as a function of x, denoted $\rho(x)$. We assume $\rho(x)$ is integrable. Because density is a function of x, we partition the interval from $[0, r]$ along the x-axis. For $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of the interval $[0, r]$, and for $i = 1, 2, \ldots, n$, choose an arbitrary point $x_i^* \in [x_{i-1}, x_i]$. Now, use the partition to break up the disk into thin (two-dimensional) washers. A disk and a representative washer are depicted in the following figure.

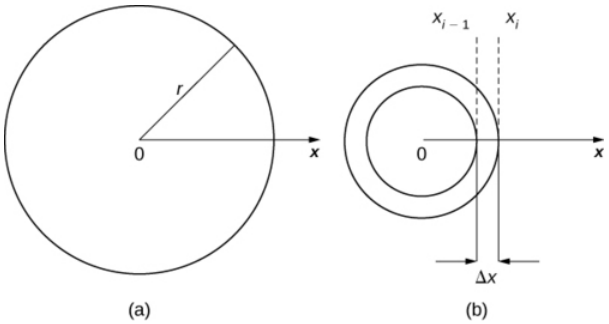


Figure 2.5.3: (a) A thin disk in the *xy*-plane. (b) A representative washer.

We now approximate the density and area of the washer to calculate an approximate mass, m_i. Note that the area of the washer is given by

$$\begin{aligned} A_i &= \pi(x_i)^2 - \pi(x_{i-1})^2 \\ &= \pi[x_i^2 - x_{i-1}^2] \\ &= \pi(x_i + x_{i-1})(x_i - x_{i-1}) \\ &= \pi(x_i + x_{i-1})\Delta x. \end{aligned}$$

You may recall that we had an expression similar to this when we were computing volumes by shells. As we did there, we use $x_i^* \approx (x_i + x_{i-1})/2$ to approximate the average radius of the washer. We obtain

$$A_i = \pi(x_i + x_{i-1})\Delta x \approx 2\pi x_i^* \Delta x.$$

Using $\rho(x_i^*)$ to approximate the density of the washer, we approximate the mass of the washer by

$$m_i \approx 2\pi x_i^* \rho(x_i^*)\Delta x.$$

Adding up the masses of the washers, we see the mass m of the entire disk is approximated by

$$m = \sum_{i=1}^{n} m_i \approx \sum_{i=1}^{n} 2\pi x_i^* \rho(x_i^*)\Delta x.$$

We again recognize this as a Riemann sum, and take the limit as $n \to \infty$. This gives us

$$\begin{aligned} m &= \lim_{n\to\infty} \sum_{i=1}^{n} 2\pi x_i^* \rho(x_i^*)\Delta x \\ &= \int_0^r 2\pi x \rho(x) dx. \end{aligned}$$

We summarize these findings in the following theorem.

Mass–Density Formula of a Circular Object

Let $\rho(x)$ be an integrable function representing the radial density of a disk of radius r. Then the mass of the disk is given by

$$m = \int_0^r 2\pi x \rho(x) dx. \tag{2.5.2}$$

Example 2.5.2: Calculating Mass from Radial Density

Let $\rho(x) = \sqrt{x}$ represent the radial density of a disk. Calculate the mass of a disk of radius 4.

Solution

Applying Equation 2.5.2, we find

$$m = \int_0^r 2\pi x \rho(x)dx$$
$$= \int_0^4 2\pi x\sqrt{x}dx = 2\pi \int_0^4 x^{3/2}dx$$
$$= 2\pi \frac{2}{5}x^{5/2}\,|_0^4 = \frac{4\pi}{5}[32]$$
$$= \frac{128\pi}{5}.$$

? Exercise 2.5.2

Let $\rho(x) = 3x + 2$ represent the radial density of a disk. Calculate the mass of a disk of radius 2.

Hint

Use the process from the previous example.

Solution

24π

Work Done by a Force

We now consider work. In physics, work is related to force, which is often intuitively defined as a push or pull on an object. When a force moves an object, we say the force does work on the object. In other words, work can be thought of as the amount of energy it takes to move an object. According to physics, when we have a constant force, work can be expressed as the product of force and distance.

In the English system, the unit of force is the pound and the unit of distance is the foot, so work is given in foot-pounds. In the metric system, kilograms and meters are used. One newton is the force needed to accelerate 1 kilogram of mass at the rate of 1 m/sec^2. Thus, the most common unit of work is the newton-meter. This same unit is also called the joule. Both are defined as kilograms times meters squared over seconds squared $(kg \cdot m^2/s^2)$.

When we have a constant force, things are pretty easy. It is rare, however, for a force to be constant. The work done to compress (or elongate) a spring, for example, varies depending on how far the spring has already been compressed (or stretched). We look at springs in more detail later in this section.

Suppose we have a variable force $F(x)$ that moves an object in a positive direction along the x-axis from point a to point b. To calculate the work done, we partition the interval $[a, b]$ and estimate the work done over each subinterval. So, for $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of the interval $[a, b]$, and for $i = 1, 2, \ldots, n$, choose an arbitrary point $x_i^* \in [x_{i-1}, x_i]$. To calculate the work done to move an object from point x_{i-1} to point x_i, we assume the force is roughly constant over the interval, and use $F(x_i^*)$ to approximate the force. The work done over the interval $[x_{i-1}, x_i]$, then, is given by

$$W_i \approx F(x_i^*)(x_i - x_{i-1}) = F(x_i^*)\Delta x.$$

Therefore, the work done over the interval $[a, b]$ is approximately

$$W = \sum_{i=1}^{n} W_i \approx \sum_{i=1}^{n} F(x_i^*)\Delta x.$$

Taking the limit of this expression as $n \to \infty$ gives us the exact value for work:

$$W = \lim_{n\to\infty} \sum_{i=1}^{n} F(x_i^*)\Delta x$$
$$= \int_a^b F(x)dx.$$

Thus, we can define work as follows.

Definition: Work

If a variable force $F(x)$ moves an object in a positive direction along the x-axis from point a to point b, then the work done on the object is

$$W = \int_a^b F(x)dx. \tag{2.5.3}$$

Note that if F is constant, the integral evaluates to $F \cdot (b-a) = F \cdot d$, which is the formula we stated at the beginning of this section.

Now let's look at the specific example of the work done to compress or elongate a spring. Consider a block attached to a horizontal spring. The block moves back and forth as the spring stretches and compresses. Although in the real world we would have to account for the force of friction between the block and the surface on which it is resting, we ignore friction here and assume the block is resting on a frictionless surface. When the spring is at its natural length (at rest), the system is said to be at equilibrium. In this state, the spring is neither elongated nor compressed, and in this equilibrium position the block does not move until some force is introduced. We orient the system such that $x = 0$ corresponds to the equilibrium position (Figure 2.5.4).

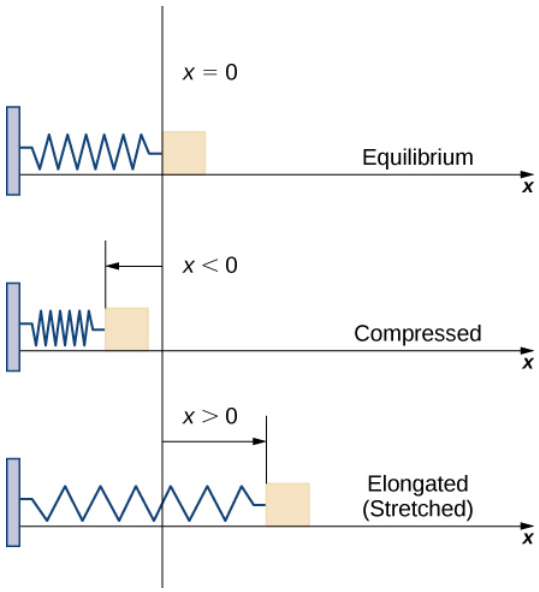


Figure 2.5.4: A block attached to a horizontal spring at equilibrium, compressed, and elongated.

According to Hooke's law, the force required to compress or stretch a spring from an equilibrium position is given by $F(x) = kx$, for some constant k. The value of k depends on the physical characteristics of the spring. The constant k is called the spring constant and is always positive. We can use this information to calculate the work done to compress or elongate a spring, as shown in the following example.

Example 2.5.3: The Work Required to Stretch or Compress a Spring

Suppose it takes a force of 10 N (in the negative direction) to compress a spring 0.2 m from the equilibrium position. How much work is done to stretch the spring 0.5 m from the equilibrium position?

Solution

First find the spring constant, k. When $x = -0.2$, we know $F(x) = -10$, so

$$\begin{aligned} F(x) &= kx \\ -10 &= k(-0.2) \\ k &= 50 \end{aligned}$$

and $F(x) = 50x$. Then, to calculate work, we integrate the force function, obtaining

$$\begin{aligned} W &= \int_a^b F(x)dx \\ &= \int_0^{0.5} 50x\, dx \\ &= 25x^2\Big|_0^{0.5} \\ &= 6.25. \end{aligned}$$

The work done to stretch the spring is 6.25 J.

? Exercise 2.5.3

Suppose it takes a force of 8 lb to stretch a spring 6 in. from the equilibrium position. How much work is done to stretch the spring 1 ft from the equilibrium position?

Hint

Use the process from the previous example. Be careful with units.

Solution

8 ft-lb

Work Done in Pumping

Consider the work done to pump water (or some other liquid) out of a tank. Pumping problems are a little more complicated than spring problems because many of the calculations depend on the shape and size of the tank. In addition, instead of being concerned about the work done to move a single mass, we are looking at the work done to move a volume of water, and it takes more work to move the water from the bottom of the tank than it does to move the water from the top of the tank.

We examine the process in the context of a cylindrical tank, then look at a couple of examples using tanks of different shapes. Assume a cylindrical tank of radius 4 m and height 10 m is filled to a depth of 8 m. How much work does it take to pump all the water over the top edge of the tank?

The first thing we need to do is define a frame of reference. We let x represent the vertical distance below the top of the tank. That is, we orient the x-axis vertically, with the origin at the top of the tank and the downward direction being positive (Figure $2.5.5$).

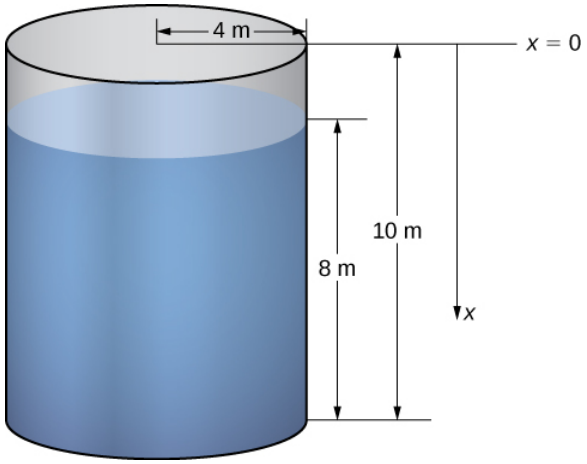


Figure $2.5.5$: How much work is needed to empty a tank partially filled with water?

Using this coordinate system, the water extends from $x = 2$ to $x = 10$. Therefore, we partition the interval $[2, 10]$ and look at the work required to lift each individual "layer" of water. So, for $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of the interval $[2, 10]$, and for $i = 1, 2, \ldots, n$, choose an arbitrary point $x_i^* \in [x_{i-1}, x_i]$. Figure $2.5.6$ shows a representative layer.

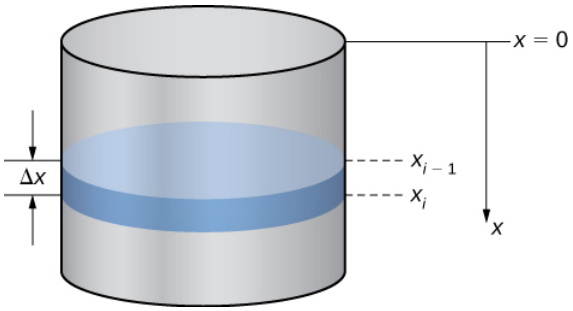


Figure 2.5.6: A representative layer of water.

In pumping problems, the force required to lift the water to the top of the tank is the force required to overcome gravity, so it is equal to the weight of the water. Given that the weight-density of water is $9800\,\mathrm{N/m^3}$, or $62.4\,\mathrm{lb/ft^3}$, calculating the volume of each layer gives us the weight. In this case, we have

$$V = \pi(4)^2 \Delta x = 16\pi\Delta x.$$

Then, the force needed to lift each layer is

$$F = 9800 \cdot 16\pi\Delta x = 156,800\pi\Delta x.$$

Note that this step becomes a little more difficult if we have a noncylindrical tank. We look at a noncylindrical tank in the next example.

We also need to know the distance the water must be lifted. Based on our choice of coordinate systems, we can use x_i^* as an approximation of the distance the layer must be lifted. Then the work to lift the ith layer of water W_i is approximately

$$W_i \approx 156,800\pi x_i^* \Delta x.$$

Adding the work for each layer, we see the approximate work to empty the tank is given by

$$W = \sum_{i=1}^{n} W_i$$

$$\approx \sum_{i=1}^{n} 156,800\pi x_i^* \Delta x.$$

This is a Riemann sum, so taking the limit as $n \to \infty$, we get

$$W = \lim_{n\to\infty} \sum_{i=1}^{n} 156,800\pi x_i^* \Delta x$$

$$= 156,800\pi \int_2^{10} x\,dx$$

$$= 156,800\pi \left(\frac{x^2}{2}\right)\Big|_2^{10} = 7,526,400\pi \approx 23,644,883.$$

The work required to empty the tank is approximately 23,650,000 J.

For pumping problems, the calculations vary depending on the shape of the tank or container. The following problem-solving strategy lays out a step-by-step process for solving pumping problems.

Problem-Solving Strategy: Solving Pumping Problems

1. Sketch a picture of the tank and select an appropriate frame of reference.
2. Calculate the volume of a representative layer of water.
3. Multiply the volume by the weight-density of water to get the force.
4. Calculate the distance the layer of water must be lifted.
5. Multiply the force and distance to get an estimate of the work needed to lift the layer of water.

6. Sum the work required to lift all the layers. This expression is an estimate of the work required to pump out the desired amount of water, and it is in the form of a Riemann sum.
7. Take the limit as $n \to \infty$ and evaluate the resulting integral to get the exact work required to pump out the desired amount of water.

We now apply this problem-solving strategy in an example with a noncylindrical tank.

Example 2.5.4: A Pumping Problem with a Noncylindrical Tank

Assume a tank in the shape of an inverted cone, with height 12 ft and base radius 4 ft. The tank is full to start with, and water is pumped over the upper edge of the tank until the height of the water remaining in the tank is 4 ft. How much work is required to pump out that amount of water?

Solution

The tank is depicted in Figure 2.5.7. As we did in the example with the cylindrical tank, we orient the x-axis vertically, with the origin at the top of the tank and the downward direction being positive (step 1).

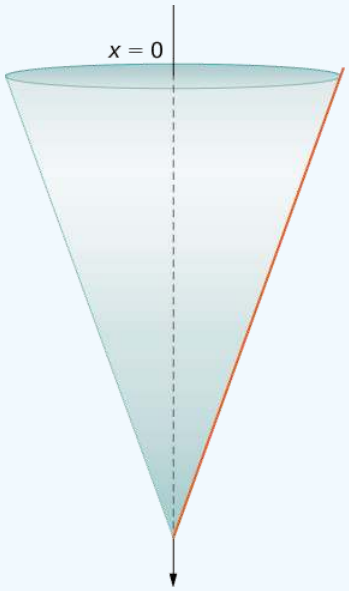


Figure 2.5.7: A water tank in the shape of an inverted cone.

The tank starts out full and ends with 4 ft of water left, so, based on our chosen frame of reference, we need to partition the interval $[0, 8]$. Then, for $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of the interval $[0, 8]$, and for $i = 1, 2, \ldots, n$, choose an arbitrary point $x_i^* \in [x_{i-1}, x_i]$. We can approximate the volume of a layer by using a disk, then use similar triangles to find the radius of the disk (Figure 2.5.8).

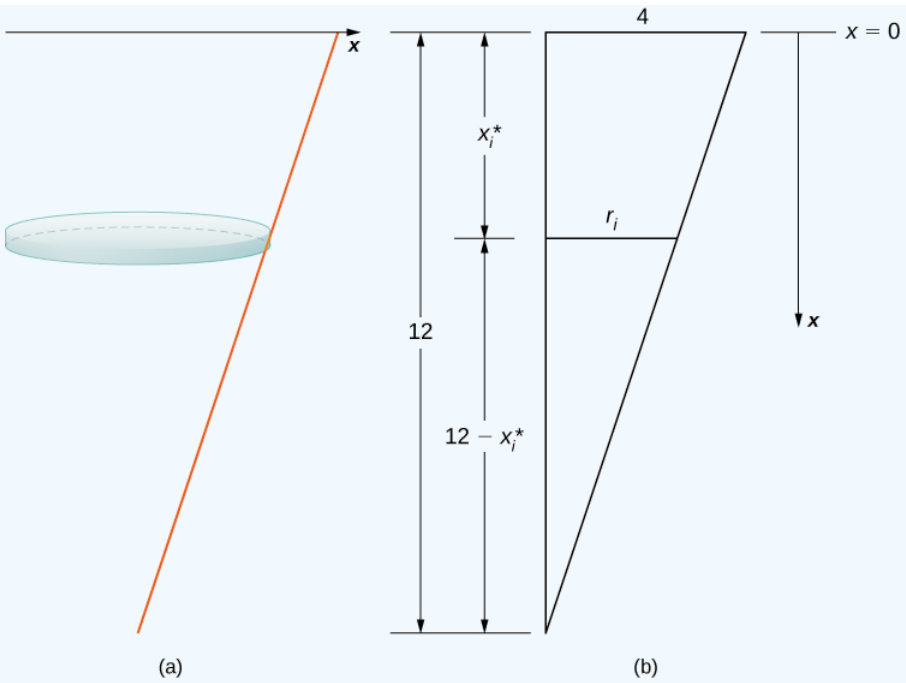


Figure 2.5.8: Using similar triangles to express the radius of a disk of water.

From properties of similar triangles, we have

$$\begin{aligned} \frac{r_i}{12-x_i^*} &= \frac{4}{12} = \frac{1}{3} && \text{(step 1)} \\ 3r_i &= 12 - x_i^* \\ r_i &= \frac{12-x_i^*}{3} \\ &= 4 - \frac{x_i^*}{3}. \end{aligned}$$

Then the volume of the disk is

$$V_i = \pi \left(4 - \frac{x_i^*}{3}\right)^2 \Delta x. \qquad \text{(step 2)}$$

The weight-density of water is 62.4lb/ft^3, so the force needed to lift each layer is approximately

$$F_i \approx 62.4\pi \left(4 - \frac{x_i^*}{3}\right)^2 \Delta x \qquad \text{(step 3)}$$

Based on the diagram, the distance the water must be lifted is approximately x_i^* feet (step 4), so the approximate work needed to lift the layer is

$$W_i \approx 62.4\pi x_i^* \left(4 - \frac{x_i^*}{3}\right)^2 \Delta x. \qquad \text{(step 5)}$$

Summing the work required to lift all the layers, we get an approximate value of the total work:

$$W = \sum_{i=1}^{n} W_i \approx \sum_{i=1}^{n} 62.4\pi x_i^* \left(4 - \frac{x_i^*}{3}\right)^2 \Delta x. \qquad \text{(step 6)}$$

Taking the limit as $n \to \infty$, we obtain

$$W = \lim_{n\to\infty} \sum_{i=1}^{n} 62.4\pi x_i^* (4 - \frac{x_i^*}{3})^2 \Delta x$$

$$= \int_0^8 62.4\pi x \left(4 - \frac{x}{3}\right)^2 dx$$

$$= 62.4\pi \int_0^8 x \left(16 - \frac{8x}{3} + \frac{x^2}{9}\right) dx = 62.4\pi \int_0^8 \left(16x - \frac{8x^2}{3} + \frac{x^3}{9}\right) dx$$

$$= 62.4\pi \left[8x^2 - \frac{8x^3}{9} + \frac{x^4}{36}\right]\Big|_0^8 = 10,649.6\pi \approx 33,456.7.$$

It takes approximately $33,450$ ft-lb of work to empty the tank to the desired level.

Exercise 2.5.4

A tank is in the shape of an inverted cone, with height 10 ft and base radius 6 ft. The tank is filled to a depth of 8 ft to start with, and water is pumped over the upper edge of the tank until 3 ft of water remain in the tank. How much work is required to pump out that amount of water?

Hint

Use the process from the previous example.

Solution

Approximately $43,255.2$ ft-lb

Hydrostatic Force and Pressure

In this last section, we look at the force and pressure exerted on an object submerged in a liquid. In the English system, force is measured in pounds. In the metric system, it is measured in newtons. Pressure is force per unit area, so in the English system we have pounds per square foot (or, perhaps more commonly, pounds per square inch, denoted psi). In the metric system we have newtons per square meter, also called **pascals.**

Let's begin with the simple case of a plate of area A submerged horizontally in water at a depth s (Figure 2.5.9). Then, the force exerted on the plate is simply the weight of the water above it, which is given by $F = \rho A s$, where ρ is the weight density of water (weight per unit volume). To find the **hydrostatic pressure**—that is, the pressure exerted by water on a submerged object—we divide the force by the area. So the pressure is $p = F/A = \rho s$.

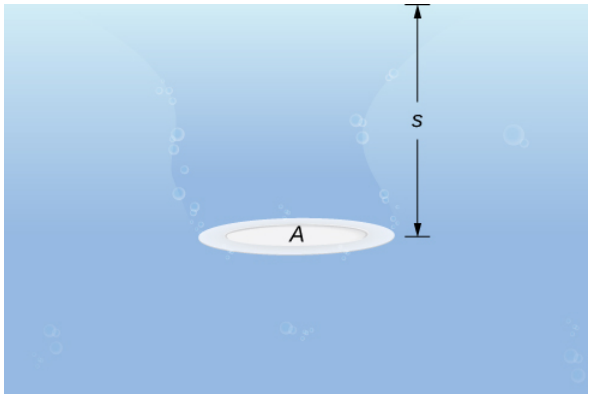


Figure 2.5.9: A plate submerged horizontally in water.

By Pascal's principle, the pressure at a given depth is the same in all directions, so it does not matter if the plate is submerged horizontally or vertically. So, as long as we know the depth, we know the pressure. We can apply Pascal's principle to find the force exerted on surfaces, such as dams, that are oriented vertically. We cannot apply the formula $F = \rho A s$ directly, because the depth varies from point to point on a vertically oriented surface. So, as we have done many times before, we form a partition, a Riemann sum, and, ultimately, a definite integral to calculate the force.

Suppose a thin plate is submerged in water. We choose our frame of reference such that the x-axis is oriented vertically, with the downward direction being positive, and point $x = 0$ corresponding to a logical reference point. Let $s(x)$ denote the depth at point x. Note we often let $x = 0$ correspond to the surface of the water. In this case, depth at any point is simply given by $s(x) = x$. However, in some cases we may want to select a different reference point for $x = 0$, so we proceed with the development in the more general case. Last, let $w(x)$ denote the width of the plate at the point x.

Assume the top edge of the plate is at point $x = a$ and the bottom edge of the plate is at point $x = b$. Then, for $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of the interval $[a, b]$, and for $i = 1, 2, \ldots, n$, choose an arbitrary point $x_i^* \in [x_{i-1}, x_i]$. The partition divides the plate into several thin, rectangular strips (Figure 2.5.10).

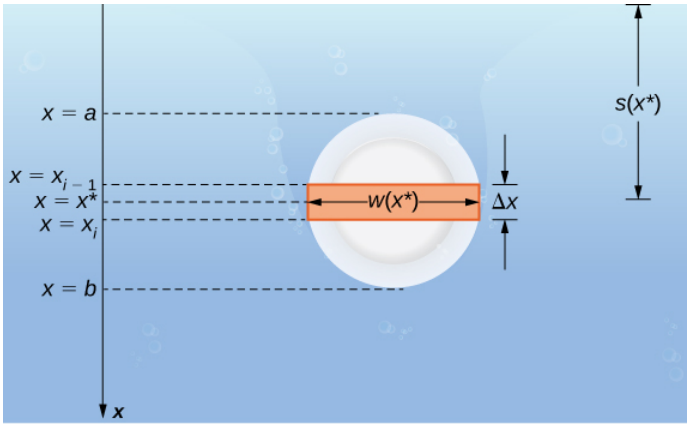


Figure 2.5.10: A thin plate submerged vertically in water.

Let's now estimate the force on a representative strip. If the strip is thin enough, we can treat it as if it is at a constant depth, $s(x_i^*)$. We then have

$$F_i = \rho A s = \rho[w(x_i^*)\Delta x]s(x_i^*).$$

Adding the forces, we get an estimate for the force on the plate:

$$F \approx \sum_{i=1}^{n} F_i = \sum_{i=1}^{n} \rho[w(x_i^*)\Delta x]s(x_i^*).$$

This is a Riemann sum, so taking the limit gives us the exact force. We obtain

$$F = \lim_{n\to\infty} \sum_{i=1}^{n} \rho[w(x_i^*)\Delta x]s(x_i^*) = \int_a^b \rho w(x)s(x)dx. \tag{2.5.4}$$

Evaluating this integral gives us the force on the plate. We summarize this in the following problem-solving strategy.

Problem-Solving Strategy: Finding Hydrostatic Force

1. Sketch a picture and select an appropriate frame of reference. (Note that if we select a frame of reference other than the one used earlier, we may have to adjust Equation 2.5.4 accordingly.)
2. Determine the depth and width functions, $s(x)$ and $w(x)$.
3. Determine the weight-density of whatever liquid with which you are working. The weight-density of water is $62.4\,\text{lb/ft}^3$, or $9800\,\text{N/m}^3$.
4. Use the equation to calculate the total force.

Example 2.5.5: Finding Hydrostatic Force

A water trough 15 ft long has ends shaped like inverted isosceles triangles, with base 8 ft and height 3 ft. Find the force on one end of the trough if the trough is full of water.

Solution

Figure 2.5.11 shows the trough and a more detailed view of one end.

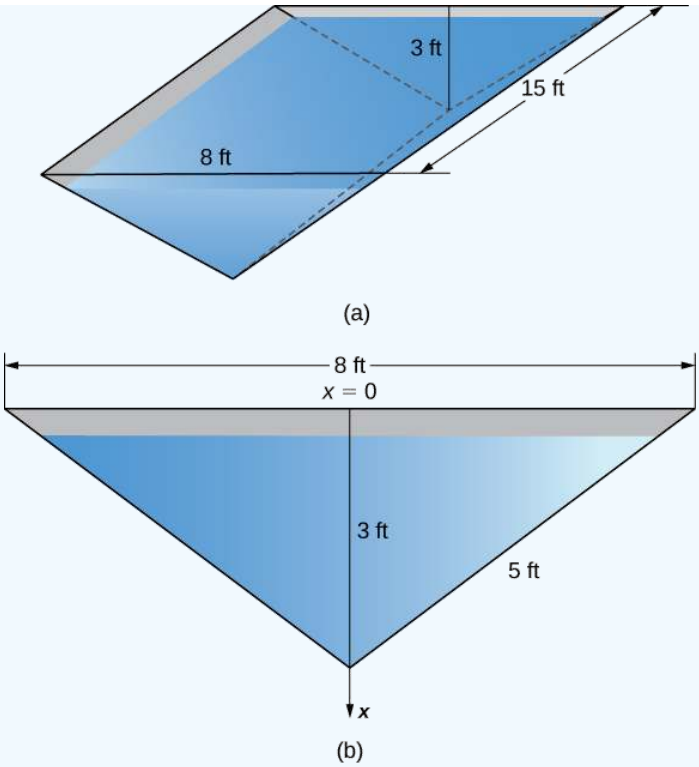


Figure 2.5.11: (a) A water trough with a triangular cross-section. (b) Dimensions of one end of the water trough.

Select a frame of reference with the x-axis oriented vertically and the downward direction being positive. Select the top of the trough as the point corresponding to $x = 0$ (step 1). The depth function, then, is $s(x) = x$. Using similar triangles, we see that $w(x) = 8 - (8/3)x$ (step 2). Now, the weight density of water is $62.4\,\text{lb/ft}^3$ (step 3), so applying Equation 2.5.4, we obtain

$$F = \int_a^b \rho w(x) s(x) dx$$

$$= \int_0^3 62.4 \left(8 - \frac{8}{3}x\right) x\, dx = 62.4 \int_0^3 \left(8x - \frac{8}{3}x^2\right) dx$$

$$= 62.4 \left[4x^2 - \frac{8}{9}x^3\right]\Big|_0^3 = 748.8.$$

The water exerts a force of 748.8 lb on the end of the trough (step 4).

Exercise 2.5.5

A water trough 12 m long has ends shaped like inverted isosceles triangles, with base 6 m and height 4 m. Find the force on one end of the trough if the trough is full of water.

Hint

Follow the problem-solving strategy and the process from the previous example.

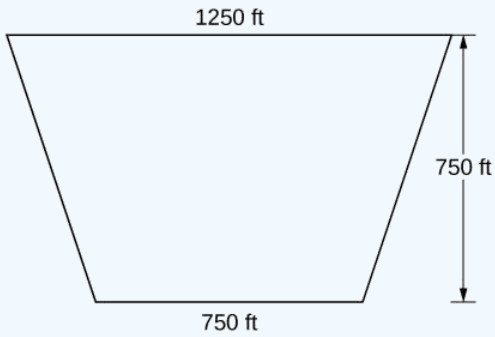
Solution

$156,800\text{N}$

Example 2.5.6: Finding Hydrostatic Force

We now return our attention to the Hoover Dam, mentioned at the beginning of this chapter. The actual dam is arched, rather than flat, but we are going to make some simplifying assumptions to help us with the calculations. Assume the face of the Hoover Dam is shaped like an isosceles trapezoid with lower base 750 ft, upper base 1250 ft, and height 750 ft (see the following figure).





When the reservoir is full, Lake Mead's maximum depth is about 530 ft, and the surface of the lake is about 10 ft below the top of the dam (see the following figure).

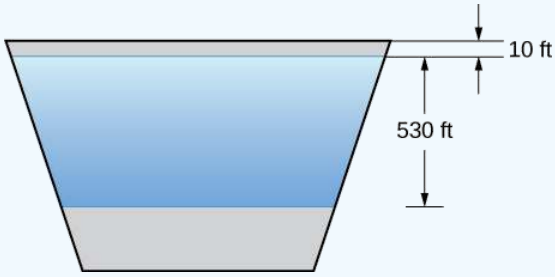


Figure 2.5.12: A simplified model of the Hoover Dam with assumed dimensions.

a. Find the force on the face of the dam when the reservoir is full.
b. The southwest United States has been experiencing a drought, and the surface of Lake Mead is about 125 ft below where it would be if the reservoir were full. What is the force on the face of the dam under these circumstances?

Solution:

a.

We begin by establishing a frame of reference. As usual, we choose to orient the x-axis vertically, with the downward direction being positive. This time, however, we are going to let $x = 0$ represent the top of the dam, rather than the surface of the water. When the reservoir is full, the surface of the water is 10 ft below the top of the dam, so $s(x) = x - 10$ (see the following figure).

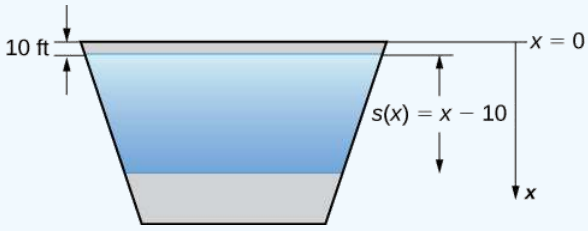


Figure 2.5.13: We first choose a frame of reference.

To find the width function, we again turn to similar triangles as shown in the figure below.

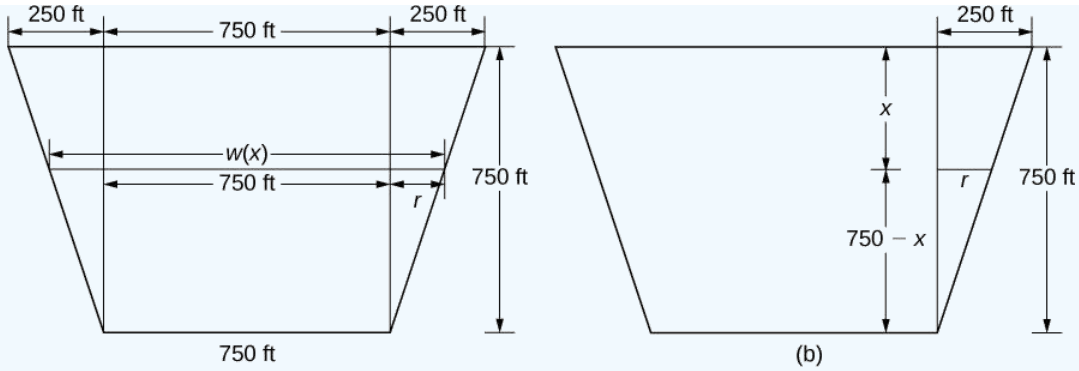


Figure 2.5.14: We use similar triangles to determine a function for the width of the dam. (a) Assumed dimensions of the dam; (b) highlighting the similar triangles.

From the figure, we see that $w(x) = 750 + 2r$. Using properties of similar triangles, we get $r = 250 - (1/3)x$. Thus,

$$w(x) = 1250 - \frac{2}{3}x \qquad \text{(step 2)}$$

Using a weight-density of 62.4lb/ft^3 (step 3) and applying Equation 2.5.4, we get

$$\begin{aligned} F &= \int_a^b \rho w(x) s(x)\, dx \\ &= \int_{10}^{540} 62.4 \left(1250 - \frac{2}{3}x\right)(x - 10)\, dx \\ &= 62.4 \int_{10}^{540} -\frac{2}{3}[x^2 - 1885x + 18750]\, dx \\ &= -62.4 \left(\frac{2}{3}\right)\left[\frac{x^3}{3} - \frac{1885x^2}{2} + 18750x\right]\Big|_{10}^{540} \approx 8,832,245,000 \text{ lb} = 4,416,122.5 \text{ t}. \end{aligned}$$

Note the change from pounds to tons (2000lb = 1 ton) (step 4). This changes our depth function, $s(x)$, and our limits of integration. We have $s(x) = x - 135$. The lower limit of integration is 135. The upper limit remains 540. Evaluating the integral, we get

$$\begin{aligned} F &= \int_a^b \rho w(x) s(x)\, dx \\ &= \int_{135}^{540} 62.4 \left(1250 - \frac{2}{3}x\right)(x - 135)\, dx \\ &= -62.4\left(\frac{2}{3}\right) \int_{135}^{540} (x - 1875)(x - 135)\, dx = -62.4 \left(\frac{2}{3}\right) \int_{135}^{540} (x^2 - 2010x + 253125)\, dx \\ &= -62.4 \left(\frac{2}{3}\right)\left[\frac{x^3}{3} - 1005x^2 + 253125x\right]\Big|_{135}^{540} \approx 5,015,230,000 \text{ lb} = 2,507,615 \text{ t}. \end{aligned}$$

Exercise 2.5.6

When the reservoir is at its average level, the surface of the water is about 50 ft below where it would be if the reservoir were full. What is the force on the face of the dam under these circumstances?

Hint

Change the depth function, $s(x)$, and the limits of integration.

Solution

Approximately 7,164,520,000 lb or 3,582,260 t

Key Concepts

- Several physical applications of the definite integral are common in engineering and physics.
- Definite integrals can be used to determine the mass of an object if its density function is known.
- Work can also be calculated from integrating a force function, or when counteracting the force of gravity, as in a pumping problem.
- Definite integrals can also be used to calculate the force exerted on an object submerged in a liquid.

Key Equations

- **Mass of a one-dimensional object**

$$m = \int_a^b \rho(x)dx$$

- **Mass of a circular object**

$$m = \int_0^r 2\pi x\rho(x)dx$$

- **Work done on an object**

$$W = \int_a^b F(x)dx$$

- **Hydrostatic force on a plate**

$$F = \int_a^b \rho w(x)s(x)dx$$

Glossary

density function
: a density function describes how mass is distributed throughout an object; it can be a linear density, expressed in terms of mass per unit length; an area density, expressed in terms of mass per unit area; or a volume density, expressed in terms of mass per unit volume; weight-density is also used to describe weight (rather than mass) per unit volume

Hooke's law
: this law states that the force required to compress (or elongate) a spring is proportional to the distance the spring has been compressed (or stretched) from equilibrium; in other words, $F = kx$, where k is a constant

hydrostatic pressure
: the pressure exerted by water on a submerged object

work
: the amount of energy it takes to move an object; in physics, when a force is constant, work is expressed as the product of force and distance

---

This page titled 2.5: Physical Applications of Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.5: Physical Applications of Integration** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.6: Moments and Centers of Mass

Learning Objectives

- Find the center of mass of objects distributed along a line.
- Locate the center of mass of a thin plate.
- Use symmetry to help locate the centroid of a thin plate.
- Apply the theorem of Pappus for volume.

In this section, we consider centers of mass (also called *centroids*, under certain conditions) and moments. The basic idea of the center of mass is the notion of a balancing point. Many of us have seen performers who spin plates on the ends of sticks. The performers try to keep several of them spinning without allowing any of them to drop. If we look at a single plate (without spinning it), there is a sweet spot on the plate where it balances perfectly on the stick. If we put the stick anywhere other than that sweet spot, the plate does not balance and it falls to the ground. (That is why performers spin the plates; the spin helps keep the plates from falling even if the stick is not exactly in the right place.) Mathematically, that sweet spot is called the *center of mass of the plate*.

In this section, we first examine these concepts in a one-dimensional context, then expand our development to consider centers of mass of two-dimensional regions and symmetry. Last, we use centroids to find the volume of certain solids by applying the theorem of Pappus.

Center of Mass and Moments

Let's begin by looking at the center of mass in a one-dimensional context. Consider a long, thin wire or rod of negligible mass resting on a fulcrum, as shown in Figure $2.6.1a$. Now suppose we place objects having masses m_1 and m_2 at distances d_1 and d_2 from the fulcrum, respectively, as shown in Figure $2.6.1b$.

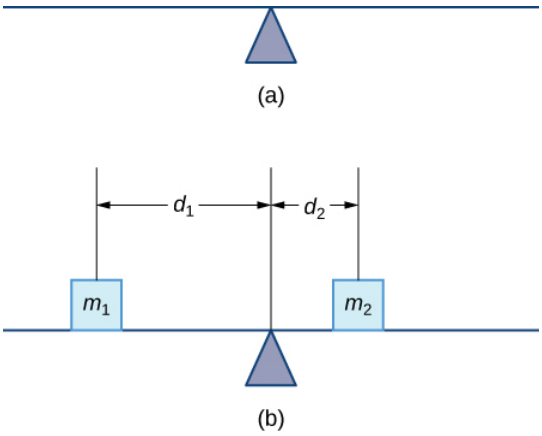


Figure 2.6.1: (a) A thin rod rests on a fulcrum. (b) Masses are placed on the rod.

The most common real-life example of a system like this is a playground seesaw, or teeter-totter, with children of different weights sitting at different distances from the center. On a seesaw, if one child sits at each end, the heavier child sinks down and the lighter child is lifted into the air. If the heavier child slides in toward the center, though, the seesaw balances. Applying this concept to the masses on the rod, we note that the masses balance each other if and only if

$$m_1 d_1 = m_2 d_2.$$

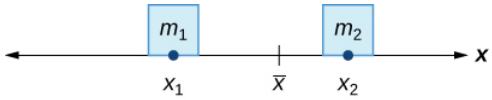


Figure 2.6.2: The center of mass $\bar{x}$ is the balance point of the system.

In the seesaw example, we balanced the system by moving the masses (children) with respect to the fulcrum. However, we are really interested in systems in which the masses are not allowed to move, and instead we balance the system by moving the

fulcrum. Suppose we have two point masses, m_1 and m_2, located on a number line at points x_1 and x_2, respectively (Figure 2.6.2). The center of mass, $\bar{x}$, is the point where the fulcrum should be placed to make the system balance.

Thus, we have

$$\begin{aligned} m_1|x_1-\bar{x}| &= m_2|x_2-\bar{x}| \\ m_1(\bar{x}-x_1) &= m_2(x_2-\bar{x}) \\ m_1\bar{x}-m_1x_1 &= m_2x_2-m_2\bar{x} \\ \bar{x}(m_1+m_2) &= m_1x_1+m_2x_2 \end{aligned}$$

or

$$\bar{x}=\frac{m_1x_1+m_2x_2}{m_1+m_2} \tag{2.6.1}$$

The expression in the numerator of Equation 2.6.1, $m_1x_1+m_2x_2$, is called the *first moment of the system* with respect to the origin. If the context is clear, we often drop the word first and just refer to this expression as the **moment** of the system. The expression in the denominator, m_1+m_2, is the total mass of the system. Thus, the **center of mass** of the system is the point at which the total mass of the system could be concentrated without changing the moment.

This idea is not limited just to two point masses. In general, if n masses, $m_1, m_2, \ldots, m_n$, are placed on a number line at points $x_1, x_2, \ldots, x_n$, respectively, then the center of mass of the system is given by

$$\bar{x}=\frac{\sum_{i=1}^{n} m_ix_i}{\sum_{i=1}^{n} m_i}$$

Center of Mass of Objects on a Line

Let $m_1, m_2, \ldots, m_n$ be point masses placed on a number line at points $x_1, x_2, \ldots, x_n$, respectively, and let $m=\sum_{i=1}^{n} m_i$ denote the total mass of the system. Then, the *moment of the system* with respect to the origin is given by

$$M=\sum_{i=1}^{n} m_ix_i \tag{2.6.2}$$

and the *center of mass of the system* is given by

$$\bar{x}=\frac{M}{m}. \tag{2.6.3}$$

We apply this theorem in the following example.

Example 2.6.1: Finding the Center of Mass of Objects along a Line

Suppose four point masses are placed on a number line as follows:

- $m_1=30\,kg,$ placed at $x_1=-2m$
- $m_2=5\,kg,$ placed at $x_2=3m$
- $m_3=10\,kg,$ placed at $x_3=6m$
- $m_4=15\,kg,$ placed at $x_4=-3m.$

Solution

Find the moment of the system with respect to the origin and find the center of mass of the system.

First, we need to calculate the moment of the system (Equation 2.6.2):

$$\begin{aligned} M &= \sum_{i=1}^{4} m_i x_i \\ &= -60 + 15 + 60 - 45 \\ &= -30. \end{aligned}$$

Now, to find the center of mass, we need the total mass of the system:

$$\begin{aligned} m &= \sum_{i=1}^{4} m_i \\ &= 30 + 5 + 10 + 15 \\ &= 60\ kg \end{aligned}$$

Then we have (from Equation 2.6.3)

$$\bar{x} - = \frac{M}{m} = -\frac{30}{60} = -\frac{1}{2}\ .$$

The center of mass is located 1/2 m to the left of the origin.

? Exercise 2.6.1

Suppose four point masses are placed on a number line as follows:

- $m_1 = 12\ kg$ placed at $x_1 = -4m$
- $m_2 = 12\ kg$ placed at $x_2 = 4m$
- $m_3 = 30\ kg$ placed at $x_3 = 2m$
- $m_4 = 6\ kg,$ placed at $x_4 = -6m.$

Find the moment of the system with respect to the origin and find the center of mass of the system.

Hint

Use the process from the previous example.

Answer

$M = 24, \bar{x} = \frac{2}{5}m$

We can generalize this concept to find the center of mass of a system of point masses in a plane. Let m_1 be a point mass located at point (x_1, y_1) in the plane. Then the moment M_x of the mass with respect to the x-axis is given by $M_x = m_1 y_1$. Similarly, the moment M_y with respect to the y-axis is given by

$$M_y = m_1 x_1.$$

Notice that the x-coordinate of the point is used to calculate the moment with respect to the y-axis, and vice versa. The reason is that the x-coordinate gives the distance from the point mass to the y-axis, and the y-coordinate gives the distance to the x-axis (see the following figure).

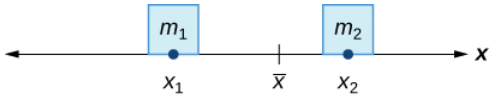


Figure 2.6.3: Point mass m_1 is located at point (x_1, y_1) in the plane.

If we have several point masses in the xy-plane, we can use the moments with respect to the x- and y-axes to calculate the x- and y-coordinates of the center of mass of the system.

Center of Mass of Objects in a Plane

Let m_1, m_2, ..., m_n be point masses located in the xy-plane at points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, respectively, and let $m = \sum_{i=1}^{n} m_i$ denote the total mass of the system. Then the moments M_x and M_y of the system with respect to the x- and y-axes, respectively, are given by

$$M_x = \sum_{i=1}^{n} m_i y_i \tag{2.6.4}$$

and

$$M_y = \sum_{i=1}^{n} m_i x_i. \tag{2.6.5}$$

Also, the coordinates of the center of mass $(\bar{x}, \bar{y})$ of the system are

$$\bar{x} = \frac{M_y}{m} \tag{2.6.6}$$

and

$$\bar{y} = \frac{M_x}{m}. \tag{2.6.7}$$

The next example demonstrates how to the center of mass formulas (Equations 2.6.4 - 2.6.7) may be applied.

Example 2.6.2: Finding the Center of Mass of Objects in a Plane

Suppose three point masses are placed in the xy-plane as follows (assume coordinates are given in meters):

- $m_1 = 2\,kg$ placed at $(-1, 3)$,
- $m_2 = 6\,kg$ placed at $(1, 1)$,
- $m_3 = 4\,kg$ placed at $(2, -2)$.

Find the center of mass of the system.

Solution

First we calculate the total mass of the system:

$$m = \sum_{i=1}^{3} m_i = 2 + 6 + 4 = 12\,kg.$$

Next we find the moments with respect to the x- and y-axes:

$$M_y = \sum_{i=1}^{3} m_i x_i = -2 + 6 + 8 = 12,$$

$$M_x = \sum_{i=1}^{3} m_i y_i = 6 + 6 - 8 = 4.$$

Then we have

$$\bar{x} = \frac{M_y}{m} = \frac{12}{12} = 1$$

and

$$\bar{y} = \frac{M_x}{m} = \frac{4}{12} = \frac{1}{3}.$$

The center of mass of the system is $(1, 1/3)$, in meters.

Exercise 2.6.2

Suppose three point masses are placed on a number line as follows (assume coordinates are given in meters):

- $m_1 = 5\,kg$, placed at $(-2, -3)$,
- $m_2 = 3\,kg$, placed at $(2, 3)$,
- $m_3 = 2\,kg$, placed at $(-3, -2)$.

Find the center of mass of the system.

Hint

Use the process from the previous example.

Answer

$(-1, -1)$ m

Center of Mass of Thin Plates

So far we have looked at systems of point masses on a line and in a plane. Now, instead of having the mass of a system concentrated at discrete points, we want to look at systems in which the mass of the system is distributed continuously across a thin sheet of material. For our purposes, we assume the sheet is thin enough that it can be treated as if it is two-dimensional. Such a sheet is called a **lamina**. Next we develop techniques to find the center of mass of a lamina. In this section, we also assume the density of the lamina is constant.

Laminas are often represented by a two-dimensional region in a plane. The geometric center of such a region is called its **centroid**. Since we have assumed the density of the lamina is constant, the center of mass of the lamina depends only on the shape of the corresponding region in the plane; it does not depend on the density. In this case, the center of mass of the lamina corresponds to the centroid of the delineated region in the plane. As with systems of point masses, we need to find the total mass of the lamina, as well as the moments of the lamina with respect to the x- and y-axes.

We first consider a lamina in the shape of a rectangle. Recall that the center of mass of a lamina is the point where the lamina balances. For a rectangle, that point is both the horizontal and vertical center of the rectangle. Based on this understanding, it is clear that the center of mass of a rectangular lamina is the point where the diagonals intersect, which is a result of the **symmetry principle**, and it is stated here without proof.

The Symmetry Principle

If a region R is symmetric about a line l, then the centroid of R lies on l.

Let's turn to more general laminas. Suppose we have a lamina bounded above by the graph of a continuous function $f(x)$, below by the x-axis, and on the left and right by the lines $x = a$ and $x = b$, respectively, as shown in the following figure.

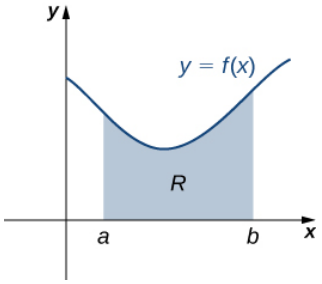


Figure 2.6.4: A region in the plane representing a lamina.

As with systems of point masses, to find the center of mass of the lamina, we need to find the total mass of the lamina, as well as the moments of the lamina with respect to the x- and y-axes. As we have done many times before, we approximate these quantities

by partitioning the interval $[a, b]$ and constructing rectangles.

For $i = 0, 1, 2, \ldots, n$, let $P = x_i$ be a regular partition of $[a, b]$. Recall that we can choose any point within the interval $[x_{i-1}, x_i]$ as our x_i^*. In this case, we want x_i^* to be the x-coordinate of the centroid of our rectangles. Thus, for $i = 1, 2, \ldots, n$, we select $x_i^* \in [x_{i-1}, x_i]$ such that x_i^* is the midpoint of the interval. That is, $x_i^* = (x_{i-1} + x_i)/2$. Now, for $i = 1, 2, \ldots, n$, construct a rectangle of height $f(x_i^*)$ on $[x_{i-1}, x_i]$. The center of mass of this rectangle is $(x_i^*, (f(x_i^*))/2)$, as shown in the following figure.

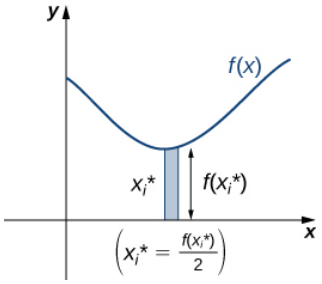


Figure 2.6.5: A representative rectangle of the lamina.

Next, we need to find the total mass of the rectangle. Let ρ represent the density of the lamina (note that ρ is a constant). In this case, ρ is expressed in terms of mass per unit area. Thus, to find the total mass of the rectangle, we multiply the area of the rectangle by ρ. Then, the mass of the rectangle is given by $\rho f(x_i^*)\Delta x$.

To get the approximate mass of the lamina, we add the masses of all the rectangles to get

$$m \approx \sum_{i=1}^{n} \rho f(x_i^*)\Delta x. \tag{2.6.8}$$

Equation 2.6.8 is a Riemann sum. Taking the limit as $n \to \infty$ gives the exact mass of the lamina:

$$\begin{aligned} m &= \lim_{n\to\infty} \sum_{i=1}^{n} \rho f(x_i^*)\Delta x \\ &= \rho \int_a^b f(x)dx. \end{aligned}$$

Next, we calculate the moment of the lamina with respect to the x-axis. Returning to the representative rectangle, recall its center of mass is $(x_i^*, (f(x_i^*))/2)$. Recall also that treating the rectangle as if it is a point mass located at the center of mass does not change the moment. Thus, the moment of the rectangle with respect to the x-axis is given by the mass of the rectangle, $\rho f(x_i^*)\Delta x$, multiplied by the distance from the center of mass to the x-axis: $(f(x_i^*))/2$. Therefore, the moment with respect to the x-axis of the rectangle is $\rho([f(x_i^*)]^2/2)\Delta x$. Adding the moments of the rectangles and taking the limit of the resulting Riemann sum, we see that the moment of the lamina with respect to the x-axis is

$$\begin{aligned} M_x &= \lim_{n\to\infty} \sum_{i=1}^{n} \rho \frac{[f(x_i^*)]^2}{2}\Delta x \\ &= \rho \int_a^b \frac{[f(x)]^2}{2}dx. \end{aligned}$$

We derive the moment with respect to the y-axis similarly, noting that the distance from the center of mass of the rectangle to the y-axis is x_i^*. Then the moment of the lamina with respect to the y-axis is given by

$$\begin{aligned} M_y &= \lim_{n\to\infty} \sum_{i=1}^{n} \rho x_i^* f(x^*)i)\Delta x \\ &= \rho \int_a^b x f(x)dx. \end{aligned}$$

We find the coordinates of the center of mass by dividing the moments by the total mass to give $\bar{x} = M_y/m$ and $\bar{y} = M_x/m$. If we look closely at the expressions for M_x, M_y, and m, we notice that the constant ρ cancels out when $\bar{x}$ and $\bar{y}$ are calculated.

We summarize these findings in the following theorem.

Center of Mass of a Thin Plate in the xy-Plane

Let R denote a region bounded above by the graph of a continuous function $f(x)$, below by the x-axis, and on the left and right by the lines $x = a$ and $x = b$, respectively. Let ρ denote the density of the associated lamina. Then we can make the following statements:

i. The mass of the lamina is

$$m = \rho \int_a^b f(x)dx. \tag{2.6.9}$$

ii. The moments M_x and M_y of the lamina with respect to the x- and y-axes, respectively, are

$$M_x = \rho \int_a^b \frac{[f(x)]^2}{2} dx \tag{2.6.10}$$

and

$$M_y = \rho \int_a^b x f(x)dx. \tag{2.6.11}$$

iii. The coordinates of the center of mass $(\bar{x}, \bar{y})$ are

$$\bar{x} = \frac{M_y}{m} \tag{2.6.12}$$

and

$$\bar{y} = \frac{M_x}{m}. \tag{2.6.13}$$

In the next example, we use this theorem to find the center of mass of a lamina.

Example 2.6.3: Finding the Center of Mass of a Lamina

Let R be the region bounded above by the graph of the function $f(x) = \sqrt{x}$ and below by the x-axis over the interval $[0, 4]$. Find the centroid of the region.

Solution

The region is depicted in the following figure.

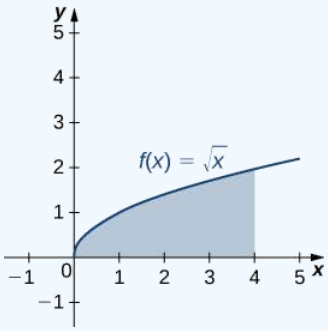


Figure 2.6.6: Finding the center of mass of a lamina.

Since we are only asked for the centroid of the region, rather than the mass or moments of the associated lamina, we know the density constant ρ cancels out of the calculations eventually. Therefore, for the sake of convenience, let's assume $\rho = 1$.

First, we need to calculate the total mass (Equation 2.6.9):

$$\begin{aligned} m &= \rho \int_a^b f(x)dx \\ &= \int_0^4 \sqrt{x}dx \\ &= \frac{2}{3}x^{3/2}\Big|_0^4 \\ &= \frac{2}{3}[8-0] \\ &= \frac{16}{3}. \end{aligned}$$

Next, we compute the moments (Equation 2.6.12):

$$\begin{aligned} M_x &= \rho \int_a^b \frac{[f(x)]^2}{2}dx \\ &= \int_0^4 \frac{x}{2}dx \\ &= \frac{1}{4}x^2\Big|_0^4 \\ &= 4 \end{aligned}$$

and (Equation 2.6.11):

$$\begin{aligned} M_y &= \rho \int_a^b xf(x)dx \\ &= \int_0^4 x\sqrt{x}dx \\ &= \int_0^4 x^{3/2}dx \\ &= \frac{2}{5}x^{5/2}\Big|_0^4 \\ &= \frac{2}{5}[32-0] \\ &= \frac{64}{5}. \end{aligned}$$

Thus, we have (Equation 2.6.12):

$$\begin{aligned} \bar{x} &= \frac{M_y}{m} \\ &= \frac{64/5}{16/3} \\ &= \frac{64}{5}\cdot\frac{3}{16} \\ &= \frac{12}{5} \end{aligned}$$

and (Equation 2.6.13):

$$\begin{aligned}\bar{y} &= \frac{M_x}{y} \\ &= \frac{4}{16/3} \\ &= 4 \cdot \frac{3}{16} \\ &= \frac{3}{4}.\end{aligned}$$

The centroid of the region is $(12/5, 3/4)$.

Exercise 2.6.3

Let R be the region bounded above by the graph of the function $f(x) = x^2$ and below by the x-axis over the interval $[0, 2]$. Find the centroid of the region.

Hint

Use the process from the previous example.

Answer

The centroid of the region is $(3/2, 6/5)$.

We can adapt this approach to find centroids of more complex regions as well. Suppose our region is bounded above by the graph of a continuous function $f(x)$, as before, but now, instead of having the lower bound for the region be the x-axis, suppose the region is bounded below by the graph of a second continuous function, $g(x)$, as shown in Figure 2.6.7.

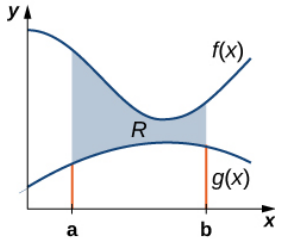


Figure 2.6.7: A region between two functions.

Again, we partition the interval $[a, b]$ and construct rectangles. A representative rectangle is shown in Figure 2.6.8.

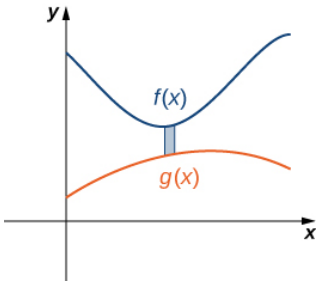


Figure 2.6.8: A representative rectangle of the region between two functions.

Note that the centroid of this rectangle is $(x_i^*, (f(x_i^*) + g(x_i^*))/2)$. We won't go through all the details of the Riemann sum development, but let's look at some of the key steps. In the development of the formulas for the mass of the lamina and the moment with respect to the y-axis, the height of each rectangle is given by $f(x_i^*) - g(x_i^*)$, which leads to the expression $f(x) - g(x)$ in the integrands.

In the development of the formula for the moment with respect to the x-axis, the moment of each rectangle is found by multiplying the area of the rectangle, $\rho[f(x_i^*) - g(x_i^*)]\Delta x$, by the distance of the centroid from the x-axis, $(f(x_i^*) + g(x_i^*))/2$, which gives $\rho(1/2)[f(x_i^*)]^2 - [g(x_i^*)]^2 \Delta x$. Summarizing these findings, we arrive at the following theorem.

Center of Mass of a Lamina Bounded by Two Functions

Let R denote a region bounded above by the graph of a continuous function $f(x)$, below by the graph of the continuous function $g(x)$, and on the left and right by the lines $x = a$ and $x = b$, respectively. Let ρ denote the density of the associated lamina. Then we can make the following statements:

i. The mass of the lamina is

$$m = \rho \int_a^b [f(x) - g(x)]dx.$$

ii. The moments M_x and M_y of the lamina with respect to the x- and y-axes, respectively, are

$$M_x = \rho \int_a^b \frac{1}{2}([f(x)]^2 - [g(x)]^2)dx$$

and

$$M_y = \rho \int_a^b x[f(x) - g(x)]dx.$$

iii. The coordinates of the center of mass $\bar{x}, \bar{y})$ are

$$\bar{x} = \frac{M_y}{m}$$

and

$$\bar{y} = \frac{M_x}{m}$$

We illustrate this theorem in the following example.

Example 2.6.4: Finding the Centroid of a Region Bounded by Two Functions

Let R be the region bounded above by the graph of the function $f(x) = 1 - x^2$ and below by the graph of the function $g(x) = x - 1$. Find the centroid of the region.

Solution

The region is depicted in the following figure.

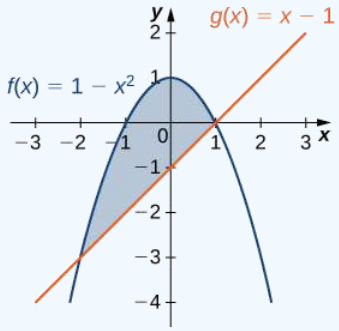


Figure 2.6.9: Finding the centroid of a region between two curves.

The graphs of the functions intersect at $(-2, -3)$ and $(1, 0)$, so we integrate from -2 to 1. Once again, for the sake of convenience, assume $\rho = 1$.

First, we need to calculate the total mass:

$$\begin{aligned} m &= \rho \int_a^b [f(x) - g(x)]dx \\ &= \int_{-2}^1 [1 - x^2 - (x-1)]dx \\ &= \int_{-2}^1 (2 - x^2 - x)dx \\ &= \left[2x - \frac{1}{3}x^3 - \frac{1}{2}x^2\right]\Big|_{-2}^1 \\ &= \left[2 - \frac{1}{3} - \frac{1}{2}\right] - \left[-4 + \frac{8}{3} - 2\right] \\ &= \frac{9}{2}. \end{aligned}$$

Next, we compute the moments:

$$\begin{aligned} M_x &= \rho \int_a^b \frac{1}{2}([f(x)]^2 - [g(x)]^2)dx \\ &= \frac{1}{2}\int_{-2}^1 ((1-x^2)^2 - (x-1)^2)dx \\ &= \frac{1}{2}\int_{-2}^1 (x^4 - 3x^2 + 2x)dx \\ &= \frac{1}{2}\left[\frac{x^5}{5} - x^3 + x^2\right]\Big|_{-2}^1 \\ &= -\frac{27}{10} \end{aligned}$$

and

$$\begin{aligned} M_y &= \rho \int_a^b x[f(x) - g(x)]dx \\ &= \int_{-2}^1 x[(1-x^2) - (x-1)]dx \\ &= \int_{-2}^1 x[2 - x^2 - x]dx \\ &= \int_{-2}^1 (2x - x^3 - x^2)dx \\ &= \left[x^2 - \frac{x^4}{4} - \frac{x^3}{3}\right]\Big|_{-2}^1 \\ &= -\frac{9}{4}. \end{aligned}$$

Therefore, we have

$$\begin{aligned}\bar{x} &= \frac{M_y}{m} \\ &= -\frac{9}{4} \cdot \frac{2}{9} \\ &= -\frac{1}{2}\end{aligned}$$

and

$$\begin{aligned}\bar{y} &= \frac{M_x}{y} \\ &= -\frac{27}{10} \cdot \frac{2}{9} \\ &= -\frac{3}{5}.\end{aligned}$$

The centroid of the region is $(-(1/2), -(3/5))$.

Exercise 2.6.4

Let R be the region bounded above by the graph of the function $f(x) = 6 - x^2$ and below by the graph of the function $g(x) = 3 - 2x$. Find the centroid of the region.

Hint

Use the process from the previous example.

Answer

The centroid of the region is $(1, 13/5)$.

The Symmetry Principle

We stated the symmetry principle earlier, when we were looking at the centroid of a rectangle. The symmetry principle can be a great help when finding centroids of regions that are symmetric. Consider the following example.

Example 2.6.5: Finding the Centroid of a Symmetric Region

Let R be the region bounded above by the graph of the function $f(x) = 4 - x^2$ and below by the x-axis. Find the centroid of the region.

Solution

The region is depicted in the following figure

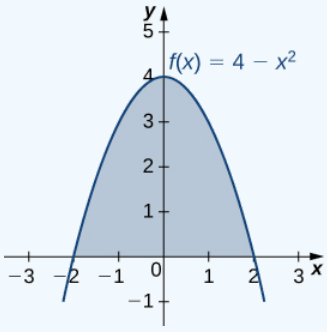


Figure 2.6.10: We can use the symmetry principle to help find the centroid of a symmetric region.

The region is symmetric with respect to the y-axis. Therefore, the x-coordinate of the centroid is zero. We need only calculate $\bar{y}$. Once again, for the sake of convenience, assume $\rho = 1$.

First, we calculate the total mass:

$$\begin{aligned} m &= \rho \int_a^b f(x)dx \\ &= \int_{-2}^{2} (4-x^2)dx \\ &= \left[4x - \frac{x^3}{3}\right]\Big|_{-2}^{2} \\ &= \frac{32}{3}. \end{aligned}$$

Next, we calculate the moments. We only need M_x:

$$\begin{aligned} M_x &= \rho \int_a^b \frac{[f(x)]^2}{2} dx \\ &= \frac{1}{2}\int_{-2}^{2} \left[4-x^2\right]^2 dx = \frac{1}{2}\int_{-2}^{2} (16 - 8x^2 + x^4)dx \\ &= \frac{1}{2}\left[\frac{x^5}{5} - \frac{8x^3}{3} + 16x\right] |_{-2}^{2} = \frac{256}{15} \end{aligned}$$

Then we have

$$\bar{y} = \frac{M_x}{y} = \frac{256}{15} \cdot \frac{3}{32} = \frac{8}{5}.$$

The centroid of the region is $(0, 8/5)$.

Exercise 2.6.5

Let R be the region bounded above by the graph of the function $f(x) = 1 - x^2$ and below by x-axis. Find the centroid of the region.

Hint

Use the process from the previous example.

Answer

The centroid of the region is $(0, 2/5)$.

The Grand Canyon Skywalk

The Grand Canyon Skywalk opened to the public on March 28, 2007. This engineering marvel is a horseshoe-shaped observation platform suspended 4000 ft above the Colorado River on the West Rim of the Grand Canyon. Its crystal-clear glass floor allows stunning views of the canyon below (see the following figure).



Figure 2.6.11: The Grand Canyon Skywalk offers magnificent views of the canyon. (credit: 10da_ralta, Wikimedia Commons)

The Skywalk is a cantilever design, meaning that the observation platform extends over the rim of the canyon, with no visible means of support below it. Despite the lack of visible support posts or struts, cantilever structures are engineered to be very stable and the Skywalk is no exception. The observation platform is attached firmly to support posts that extend 46 ft down into bedrock. The structure was built to withstand 100-mph winds and an 8.0-magnitude earthquake within 50 mi, and is capable of supporting more than 70,000,000 lb.

One factor affecting the stability of the Skywalk is the center of gravity of the structure. We are going to calculate the center of gravity of the Skywalk, and examine how the center of gravity changes when tourists walk out onto the observation platform.

The observation platform is U-shaped. The legs of the U are 10 ft wide and begin on land, under the visitors' center, 48 ft from the edge of the canyon. The platform extends 70 ft over the edge of the canyon.

To calculate the center of mass of the structure, we treat it as a lamina and use a two-dimensional region in the xy-plane to represent the platform. We begin by dividing the region into three subregions so we can consider each subregion separately. The first region, denoted R_1, consists of the curved part of the U. We model R_1 as a semicircular annulus, with inner radius 25 ft and outer radius 35 ft, centered at the origin (Figure 2.6.12).

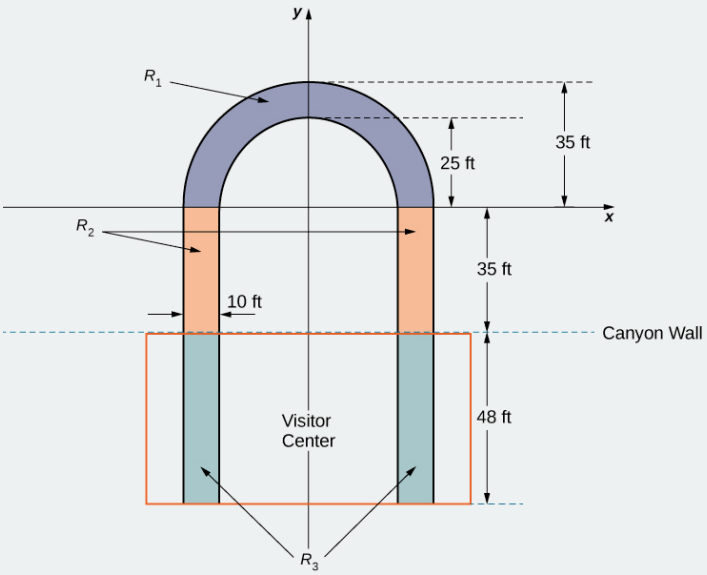


Figure 2.6.12: We model the Skywalk with three sub-regions.

The legs of the platform, extending 35 ft between R_1 and the canyon wall, comprise the second sub-region, R_2. Last, the ends of the legs, which extend 48 ft under the visitor center, comprise the third sub-region, R_3. Assume the density of the lamina is constant and assume the total weight of the platform is 1,200,000 lb (not including the weight of the visitor center; we will consider that later). Use $g = 32\ ft/sec^2$.

1. Compute the area of each of the three sub-regions. Note that the areas of regions R_2 and R_3 should include the areas of the legs only, not the open space between them. Round answers to the nearest square foot.

2. Determine the mass associated with each of the three sub-regions.
3. Calculate the center of mass of each of the three sub-regions.
4. Now, treat each of the three sub-regions as a point mass located at the center of mass of the corresponding sub-region. Using this representation, calculate the center of mass of the entire platform.
5. Assume the visitor center weighs 2,200,000 lb, with a center of mass corresponding to the center of mass of R_3.Treating the visitor center as a point mass, recalculate the center of mass of the system. How does the center of mass change?
6. Although the Skywalk was built to limit the number of people on the observation platform to 120, the platform is capable of supporting up to 800 people weighing 200 lb each. If all 800 people were allowed on the platform, and all of them went to the farthest end of the platform, how would the center of gravity of the system be affected? (Include the visitor center in the calculations and represent the people by a point mass located at the farthest edge of the platform, 70 ft from the canyon wall.)

Theorem of Pappus

This section ends with a discussion of the *theorem of Pappus for volume*, which allows us to find the volume of particular kinds of solids by using the centroid. (There is also a theorem of Pappus for surface area, but it is much less useful than the theorem for volume.)

Theorem of Pappus for Volume

Let R be a region in the plane and let l be a line in the plane that does not intersect R. Then the volume of the solid of revolution formed by revolving R around l is equal to the area of R multiplied by the distance d traveled by the centroid of R.

Proof

We can prove the case when the region is bounded above by the graph of a function $f(x)$ and below by the graph of a function $g(x)$ over an interval $[a,b]$, and for which the axis of revolution is the y-axis. In this case, the area of the region is $A=\int_a^b [f(x)-g(x)]\,dx$. Since the axis of rotation is the y-axis, the distance traveled by the centroid of the region depends only on the x-coordinate of the centroid, $\bar{x}$, which is

$$x=\frac{M_y}{m},$$

where

$$m=\rho\int_a^b [f(x)-g(x)]dx$$

and

$$M_y=\rho\int_a^b x[f(x)-g(x)]dx.$$

Then,

$$d=2\pi\frac{\rho\int_a^b x[f(x)-g(x)]dx}{\rho\int_a^b [f(x)-g(x)]dx}$$

and thus

$$d\cdot A=2\pi\int_a^b x[f(x)-g(x)]dx.$$

However, using the method of cylindrical shells, we have

$$V = 2\pi \int_a^b x[f(x) - g(x)]dx.$$

So,

$$V = d \cdot A$$

and the proof is complete.

□

Example 2.6.6: Using the Theorem of Pappus for Volume

Let R be a circle of radius 2 centered at $(4, 0)$. Use the theorem of Pappus for volume to find the volume of the torus generated by revolving R around the y-axis.

Solution

The region and torus are depicted in the following figure.

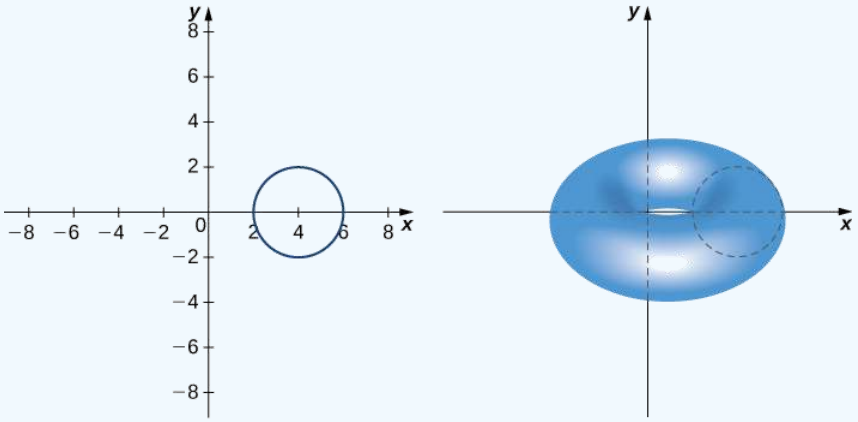


Figure 2.6.13: Determining the volume of a torus by using the theorem of Pappus. (a) A circular region R in the plane; (b) the torus generated by revolving R about the y-axis.

The region R is a circle of radius 2, so the area of R is $A = 4\pi$ units^2. By the symmetry principle, the centroid of R is the center of the circle. The centroid travels around the y-axis in a circular path of radius 4, so the centroid travels $d = 8\pi$ units. Then, the volume of the torus is $A \cdot d = 32\pi^2$ units^3.

Exercise 2.6.6

Let R be a circle of radius 1 centered at $(3, 0)$. Use the theorem of Pappus for volume to find the volume of the torus generated by revolving R around the y-axis.

Hint

Use the process from the previous example.

Answer

$6\pi^2$ units^3

Key Concepts

- Mathematically, the center of mass of a system is the point at which the total mass of the system could be concentrated without changing the moment. Loosely speaking, the center of mass can be thought of as the balancing point of the system.

- For point masses distributed along a number line, the moment of the system with respect to the origin is $M = \sum_{i=1}^{n} m_i x_i$. For point masses distributed in a plane, the moments of the system with respect to the x- and y-axes, respectively, are $M_x = \sum_{i=1}^{n} m_i y_i$ and $M_y = \sum_{i=}^{n} m_i x_i$, respectively.
- For a lamina bounded above by a function $f(x)$, the moments of the system with respect to the x- and y-axes, respectively, are $M_x = \rho \int_a^b \frac{[f(x)]^2}{2}\, dx$ and $M_y = \rho \int_a^b x f(x)\, dx$.
- The x- and y-coordinates of the center of mass can be found by dividing the moments around the y-axis and around the x-axis, respectively, by the total mass. The symmetry principle says that if a region is symmetric with respect to a line, then the centroid of the region lies on the line.
- The theorem of Pappus for volume says that if a region is revolved around an external axis, the volume of the resulting solid is equal to the area of the region multiplied by the distance traveled by the centroid of the region.

Key Equations

- **Mass of a lamina**

$$m = \rho \int_a^b f(x) dx$$

- **Moments of a lamina**

$$M_x = \rho \int_a^b \frac{[f(x)]^2}{2}\, dx \text{ and } M_y = \rho \int_a^b x f(x)\, dx$$

- **Center of mass of a lamina**

$$\bar{x} = \frac{M_y}{m} \text{ and } \bar{y} = \frac{M_x}{m}$$

Glossary

center of mass

the point at which the total mass of the system could be concentrated without changing the moment

centroid

the centroid of a region is the geometric center of the region; laminas are often represented by regions in the plane; if the lamina has a constant density, the center of mass of the lamina depends only on the shape of the corresponding planar region; in this case, the center of mass of the lamina corresponds to the centroid of the representative region

lamina

a thin sheet of material; laminas are thin enough that, for mathematical purposes, they can be treated as if they are two-dimensional

moment

if n masses are arranged on a number line, the moment of the system with respect to the origin is given by $M = \sum_{i=1}^{n} m_i x_i$; if, instead, we consider a region in the plane, bounded above by a function $f(x)$ over an interval $[a, b]$, then the moments of the region with respect to the x- and y-axes are given by $M_x = \rho \int_a^b \frac{[f(x)]^2}{2}\, dx$ and $M_y = \rho \int_a^b x f(x)\, dx$, respectively

symmetry principle

the symmetry principle states that if a region R is symmetric about a line I, then the centroid of R lies on I

theorem of Pappus for volume

this theorem states that the volume of a solid of revolution formed by revolving a region around an external axis is equal to the area of the region multiplied by the distance traveled by the centroid of the region

---

This page titled 2.6: Moments and Centers of Mass is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.6: Moments and Centers of Mass** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.7: Integrals, Exponential Functions, and Logarithms

Learning Objectives

- Write the definition of the natural logarithm as an integral.
- Recognize the derivative of the natural logarithm.
- Integrate functions involving the natural logarithmic function.
- Define the number e through an integral.
- Recognize the derivative and integral of the exponential function.
- Prove properties of logarithms and exponential functions using integrals.
- Express general logarithmic and exponential functions in terms of natural logarithms and exponentials.

We already examined exponential functions and logarithms in earlier chapters. However, we glossed over some key details in the previous discussions. For example, we did not study how to treat exponential functions with exponents that are irrational. The definition of the number e is another area where the previous development was somewhat incomplete. We now have the tools to deal with these concepts in a more mathematically rigorous way, and we do so in this section.

For purposes of this section, assume we have not yet defined the natural logarithm, the number e, or any of the integration and differentiation formulas associated with these functions. By the end of the section, we will have studied these concepts in a mathematically rigorous way (and we will see they are consistent with the concepts we learned earlier). We begin the section by defining the natural logarithm in terms of an integral. This definition forms the foundation for the section. From this definition, we derive differentiation formulas, define the number e, and expand these concepts to logarithms and exponential functions of any base.

The Natural Logarithm as an Integral

Recall the power rule for integrals:

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq -1.$$

Clearly, this does not work when $n = -1$, as it would force us to divide by zero. So, what do we do with $\int \frac{1}{x}\,dx$? Recall from the Fundamental Theorem of Calculus that $\int_1^x \frac{1}{t}dt$ is an antiderivative of $\frac{1}{x}$. Therefore, we can make the following definition.

Definition: The Natural Logarithm

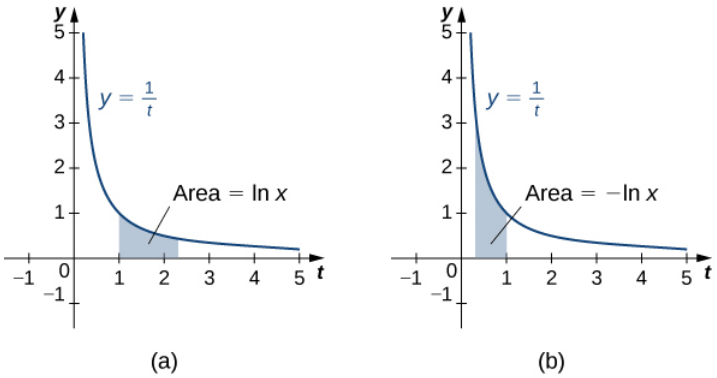
For $x > 0$, define the natural logarithm function by

$$\ln x = \int_1^x \frac{1}{t}\,dt.$$

For $x > 1$, this is just the area under the curve $y = \frac{1}{t}$ from 1 to x. For $x < 1$, we have

$$\int_1^x \frac{1}{t}\,dt = -\int_x^1 \frac{1}{t}\,dt,$$

so in this case it is the negative of the area under the curve from x to 1 (see the following figure).



(a) (b)

Figure 2.7.1: (a) When $x > 1$, the natural logarithm is the area under the curve $y = 1/t$ from 1 to x. (b) When $x < 1$, the natural logarithm is the negative of the area under the curve from x to 1.

Notice that $\ln 1 = 0$. Furthermore, the function $y = \dfrac{1}{t} > 0$ for $x > 0$. Therefore, by the properties of integrals, it is clear that $\ln x$ is increasing for $x > 0$.

Properties of the Natural Logarithm

Because of the way we defined the natural logarithm, the following differentiation formula falls out immediately as a result of to the Fundamental Theorem of Calculus.

Definition: Derivative of the Natural Logarithm

For $x > 0$, the derivative of the natural logarithm is given by

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

Corollary to the Derivative of the Natural Logarithm

The function $\ln x$ is differentiable; therefore, it is continuous.

A graph of $\ln x$ is shown in Figure. Notice that it is continuous throughout its domain of $(0, \infty)$.

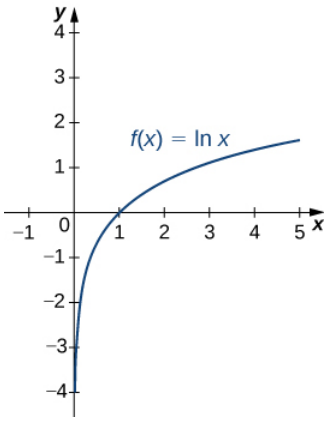


Figure 2.7.2: The graph of $f(x) = \ln x$ shows that it is a continuous function.

Example 2.7.1: Calculating Derivatives of Natural Logarithms

Calculate the following derivatives:

a. $\dfrac{d}{dx}\Big(\ln(5x^3 - 2)\Big)$

b. $\dfrac{d}{dx}\Big((\ln(3x))^2\Big)$

Solution

We need to apply the chain rule in both cases.

a. $\dfrac{d}{dx}\Big(\ln(5x^3-2)\Big) = \dfrac{15x^2}{5x^3-2}$

b. $\dfrac{d}{dx}\Big((\ln(3x))^2\Big) = \dfrac{2(\ln(3x))\cdot 3}{3x} = \dfrac{2(\ln(3x))}{x}$

? Exercise 2.7.1

Calculate the following derivatives:

a. $\dfrac{d}{dx}\Big(\ln(2x^2+x)\Big)$

b. $\dfrac{d}{dx}\Big((\ln(x^3))^2\Big)$

Hint

Apply the differentiation formula just provided and use the chain rule as necessary.

Answer

a. $\dfrac{d}{dx}\Big(\ln(2x^2+x)\Big) = \dfrac{4x+1}{2x^2+x}$

b. $\dfrac{d}{dx}\Big((\ln(x^3))^2\Big) = \dfrac{6\ln(x^3)}{x}$

Note that if we use the absolute value function and create a new function $\ln|x|$, we can extend the domain of the natural logarithm to include $x<0$. Then $\dfrac{d}{dx}\Big(\ln x\Big) = \dfrac{1}{x}$. This gives rise to the familiar integration formula.

Integral of $\frac{1}{u}\,du$

The natural logarithm is the antiderivative of the function $f(u) = \dfrac{1}{u}$:

$$\int \frac{1}{u}\,du = \ln|u| + C.$$

✓ Example 2.7.2: Calculating Integrals Involving Natural Logarithms

Calculate the integral $\displaystyle\int \frac{x}{x^2+4}\,dx$.

Solution

Using u-substitution, let $u = x^2+4$. Then $du = 2x\,dx$ and we have

$$\int \frac{x}{x^2+4}\,dx = \frac{1}{2}\int \frac{1}{u}\,du = \frac{1}{2}\ln|u| + C = \frac{1}{2}\ln|x^2+4| + C = \frac{1}{2}\ln(x^2+4) + C.$$

? Exercise 2.7.2

Calculate the integral $\displaystyle\int \frac{x^2}{x^3+6}\,dx$.

Hint

Apply the integration formula provided earlier and use u-substitution as necessary.

Answer

$$\int \frac{x^2}{x^3+6}\,dx = \frac{1}{3}\ln\,|\,x^3+6\,|\,+C$$

Although we have called our function a "logarithm," we have not actually proved that any of the properties of logarithms hold for this function. We do so here.

Properties of the Natural Logarithm

If $a,\ b>0$ and r is a rational number, then

i. $\ln 1 = 0$
ii. $\ln(ab) = \ln a + \ln b$
iii. $\ln\left(\frac{a}{b}\right) = \ln a - \ln b$
iv. $\ln(a^r) = r\ln a$

Proof

i. By definition, $\ln 1 = \int_1^1 \frac{1}{t}\,dt = 0.$

ii. We have

$$\ln(ab) = \int_1^{ab} \frac{1}{t}\,dt = \int_1^a \frac{1}{t}\,dt + \int_a^{ab} \frac{1}{t}\,dt.$$

Use $u-substitution$ on the last integral in this expression. Let $u = t/a$. Then $du = (1/a)dt$. Furthermore, when $t = a,\ u = 1$, and when $t = ab,\ u = b$. So we get

$$\ln(ab) = \int_1^a \frac{1}{t}\,dt + \int_a^{ab} \frac{1}{t}\,dt = \int_1^a \frac{1}{t}\,dt + \int_1^{ab} \frac{a}{t}\cdot\frac{1}{a}\,dt = \int_1^a \frac{1}{t}\,dt + \int_1^b \frac{1}{u}\,du = \ln a + \ln b.$$

iii. Note that

$$\frac{d}{dx}\left(\ln(x^r)\right) = \frac{rx^{r-1}}{x^r} = \frac{r}{x}\ .$$

Furthermore,

$$\frac{d}{dx}\left((r\ln x)\right) = \frac{r}{x}.$$

Since the derivatives of these two functions are the same, by the Fundamental Theorem of Calculus, they must differ by a constant. So we have

$$\ln(x^r) = r\ln x + C$$

for some constant C. Taking $x = 1$, we get

$$\ln(1^r) = r\ln(1) + C$$
$$0 = r(0) + C$$
$$C = 0.$$

Thus $\ln(x^r) = r\ln x$ and the proof is complete. Note that we can extend this property to irrational values of r later in this section.

Part iii. follows from parts ii. and iv. and the proof is left to you.

□

Example 2.7.3: Using Properties of Logarithms

Use properties of logarithms to simplify the following expression into a single logarithm:

$$\ln 9 - 2\ln 3 + \ln\left(\tfrac{1}{3}\right).$$

Solution

We have

$$\ln 9 - 2\ln 3 + \ln\left(\tfrac{1}{3}\right) = \ln(3^2) - 2\ln 3 + \ln(3^{-1}) = 2\ln 3 - 2\ln 3 - \ln 3 = -\ln 3.$$

Exercise 2.7.3

Use properties of logarithms to simplify the following expression into a single logarithm:

$$\ln 8 - \ln 2 - \ln\left(\tfrac{1}{4}\right)$$

Hint

Apply the properties of logarithms.

Answer

$4\ln 2$

Defining the Number e

Now that we have the natural logarithm defined, we can use that function to define the number e.

Definition: e

The number e is defined to be the real number such that

$$\ln e = 1$$

To put it another way, the area under the curve $y = 1/t$ between $t = 1$ and $t = e$ is 1 (Figure). The proof that such a number exists and is unique is left to you. (Hint: Use the Intermediate Value Theorem to prove existence and the fact that $\ln x$ is increasing to prove uniqueness.)

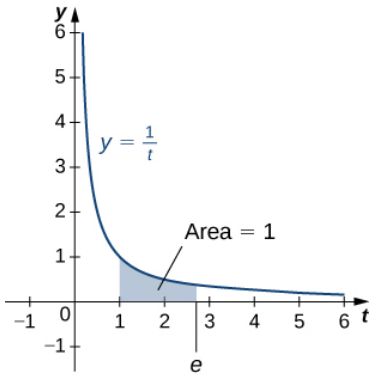


Figure 2.7.3:The area under the curve from 1 to e is equal to one.

The number e can be shown to be *irrational*, although we won't do so here (see the Student Project in Taylor and Maclaurin Series). Its approximate value is given by

$$e \approx 2.71828182846.$$

The Exponential Function

We now turn our attention to the function e^x. Note that the natural logarithm is one-to-one and therefore has an inverse function. For now, we denote this inverse function by $\exp x$. Then,

$$\exp(\ln x) = x$$

for $x > 0$ and

$$\ln(\exp x) = x$$

for all x.

The following figure shows the graphs of $\exp x$ and $\ln x$.

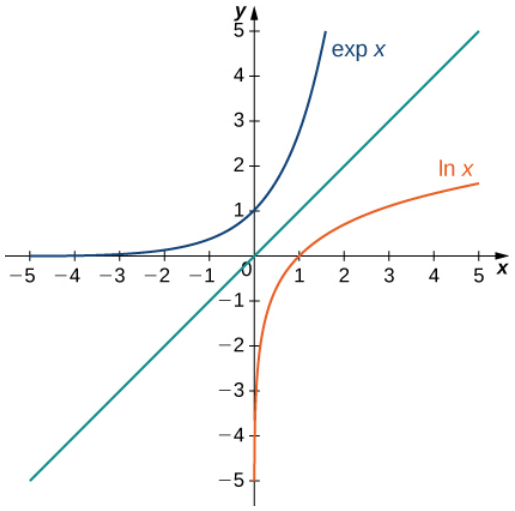


Figure 2.7.4: The graphs of $\ln x$ and $\exp x$.

We hypothesize that $\exp x = e^x$. For rational values of x, this is easy to show. If x is rational, then we have $\ln(e^x) = x \ln e = x$. Thus, when x is rational, $e^x = \exp x$. For irrational values of x, we simply define e^x as the inverse function of $\ln x$.

Definition

For any real number x, define $y = e^x$ to be the number for which

$$\ln y = \ln(e^x) = x.$$

Then we have $e^x = \exp x$ for all x, and thus

$$e^{\ln x} = x \text{ for } x > 0 \text{ and } \ln(e^x) = x$$

for all x.

Properties of the Exponential Function

Since the exponential function was defined in terms of an inverse function, and not in terms of a power of e we must verify that the usual laws of exponents hold for the function e^x.

Properties of the Exponential Function

If p and q are any real numbers and r is a rational number, then

i. $e^p e^q = e^{p+q}$
ii. $\dfrac{e^p}{e^q} = e^{p-q}$
iii. $(e^p)^r = e^{pr}$

Proof

Note that if p and q are rational, the properties hold. However, if p or q are irrational, we must apply the inverse function definition of e^x and verify the properties. Only the first property is verified here; the other two are left to you. We have

$$\ln(e^p e^q) = \ln(e^p) + \ln(eq) = p + q = \ln(e^{p+q}).$$

Since $\ln x$ is one-to-one, then

$$e^p e^q = e^{p+q}.$$

□

As with part iv. of the logarithm properties, we can extend property iii. to irrational values of r, and we do so by the end of the section.

We also want to verify the differentiation formula for the function $y = e^x$. To do this, we need to use implicit differentiation. Let $y = e^x$. Then

$$\begin{aligned} \ln y &= x \\ \frac{d}{dx}\left(\ln y\right) &= \frac{d}{dx}\left(x\right) \\ \frac{1}{y}\frac{dy}{dx} &= 1 \\ \frac{dy}{dx} &= y. \end{aligned}$$

Thus, we see

$$\frac{d}{dx}\left(e^x\right) = e^x$$

as desired, which leads immediately to the integration formula

$$\int e^x \, dx = e^x + C.$$

We apply these formulas in the following examples.

Example 2.7.4: Using Properties of Exponential Functions

Evaluate the following derivatives:

a. $\frac{d}{dt}\left(e^{3t}e^{t^2}\right)$

b. $\frac{d}{dx}\left(e^{3x^2}\right)$

Solution

We apply the chain rule as necessary.

a. $\frac{d}{dt}\left(e^{3t}e^{t^2}\right) = \frac{d}{dt}\left(e^{3t+t^2}\right) = e^{3t+t^2}(3+2t)$

b. $\frac{d}{dx}\left(e^{3x^2}\right) = e^{3x^2}6x$

Exercise 2.7.4

Evaluate the following derivatives:

a. $\frac{d}{dx}\left(\frac{e^{x^2}}{e^{5x}}\right)$

b. $\dfrac{d}{dt}\left((e^{2t})^3\right)$

Hint

Use the properties of exponential functions and the chain rule as necessary.

Answer

a. $\dfrac{d}{dx}\left(\dfrac{e^{x^2}}{e^{5x}}\right) = e^{x^{2-5x}}(2x-5)$

b. $\dfrac{d}{dt}\left((e^{2t})^3\right) = 6e^{6t}$

Example 2.7.5: Using Properties of Exponential Functions

Evaluate the following integral: $\displaystyle\int 2xe^{-x^2}\,dx$.

Solution

Using u-substitution, let $u = -x^2$. Then $du = -2x\,dx$, and we have

$$\int 2xe^{-x^2}\,dx = -\int e^u\,du = -e^u + C = -e^{-x^2} + C.$$

Exercise 2.7.5

Evaluate the following integral: $\displaystyle\int \frac{4}{e^{3x}}\,dx$.

Hint

Use the properties of exponential functions and $u-substitution$ as necessary.

Answer

$$\int \frac{4}{e^{3x}}\,dx = -\frac{4}{3}e^{-3x} + C$$

General Logarithmic and Exponential Functions

We close this section by looking at exponential functions and logarithms with bases other than e. Exponential functions are functions of the form $f(x) = a^x$. Note that unless $a = e$, we still do not have a mathematically rigorous definition of these functions for irrational exponents. Let's rectify that here by defining the function $f(x) = a^x$ in terms of the exponential function e^x. We then examine logarithms with bases other than e as inverse functions of exponential functions.

Definition: Exponential Function

For any $a > 0$, and for any real number x, define $y = a^x$ as follows:

$$y = a^x = e^{x\ln a}.$$

Now a^x is defined rigorously for all values of x. This definition also allows us to generalize property iv. of logarithms and property iii. of exponential functions to apply to both rational and irrational values of r. It is straightforward to show that properties of exponents hold for general exponential functions defined in this way.

Let's now apply this definition to calculate a differentiation formula for a^x. We have

$$\frac{d}{dx}\left(a^x\right) = \frac{d}{dx}\left(e^{x\ln a}\right) = e^{x\ln a}\ln a = a^x \ln a.$$

The corresponding integration formula follows immediately.

Derivatives and Integrals Involving General Exponential Functions

Let $a > 0$. Then,

$$\frac{d}{dx}\left(a^x\right) = a^x \ln a$$

and

$$\int a^x \, dx = \frac{1}{\ln a} a^x + C.$$

If $a \neq 1$, then the function a^x is one-to-one and has a well-defined inverse. Its inverse is denoted by $\log_a x$. Then,

$$y = \log_a x \text{ if and only if } x = a^y.$$

Note that general logarithm functions can be written in terms of the natural logarithm. Let $y = \log_a x$. Then, $x = a^y$. Taking the natural logarithm of both sides of this second equation, we get

$$\begin{aligned} \ln x &= \ln(a^y) \\ \ln x &= y \ln a \\ y &= \frac{\ln x}{\ln a} \\ \log_a x &= \frac{\ln x}{\ln a}. \end{aligned}$$

Thus, we see that all logarithmic functions are constant multiples of one another. Next, we use this formula to find a differentiation formula for a logarithm with base a. Again, let $y = \log_a x$. Then,

$$\begin{aligned} \frac{dy}{dx} &= \frac{d}{dx}\left(\log_a x\right) \\ &= \frac{d}{dx}\left(\frac{\ln x}{\ln a}\right) \\ &= (\frac{1}{\ln a})\frac{d}{dx}\left(\ln x\right) \\ &= \frac{1}{\ln a} \cdot \frac{1}{x} = \frac{1}{x \ln a} \end{aligned}$$

Derivatives of General Logarithm Functions

Let $a > 0$. Then,

$$\frac{d}{dx}\left(\log_a x\right) = \frac{1}{x \ln a}.$$

Example 2.7.6: Calculating Derivatives of General Exponential and Logarithm Functions

Evaluate the following derivatives:

a. $\frac{d}{dt}\left(4^t \cdot 2^{t^2}\right)$

b. $\frac{d}{dx}\left(\log_8(7x^2+4)\right)$

Solution: We need to apply the chain rule as necessary.

a. $\frac{d}{dt}\left(4^t \cdot 2^{t^2}\right) = \frac{d}{dt}\left(2^{2t} \cdot 2^{t^2}\right) = \frac{d}{dt}\left(2^{2t+t^2}\right) = 2^{2t+t^2} \ln(2)(2+2t)$

b. $\frac{d}{dx}\left(\log_8(7x^2+4)\right) = \frac{1}{(7x^2+4)(\ln 8)}(14x)$

? Exercise 2.7.6

Evaluate the following derivatives:

a. $\frac{d}{dt}\left(4^{t^4}\right)$

b. $\frac{d}{dx}\left(\log_3(\sqrt{x^2+1})\right)$

Hint

Use the formulas and apply the chain rule as necessary.

Answer

a. $\frac{d}{dt}\left(4^{t^4}\right) = 4^{t^4}(\ln 4)(4t^3)$

b. $\frac{d}{dx}\left(\log_3(\sqrt{x^2+1})\right) = \frac{x}{(\ln 3)(x^2+1)}$

✓ Example 2.7.7: Integrating General Exponential Functions

Evaluate the following integral: $\int \frac{3}{2^{3x}}\,dx$.

Solution

Use $u-substitution$ and let $u=-3x$. Then $du=-3\,dx$ and we have

$$\int \frac{3}{2^{3x}}\,dx = \int 3\cdot 2^{-3x}\,dx = -\int 2^u\,du = -\frac{1}{\ln 2}2^u + C = -\frac{1}{\ln 2}2^{-3x} + C.$$

? Exercise 2.7.7

Evaluate the following integral: $\int x^2 2^{x^3}\,dx$.

Hint

Use the properties of exponential functions and u-substitution

Answer

$$\int x^2 2^{x^3}\,dx = \frac{1}{3\ln 2}2^{x^3} + C$$

Key Concepts

- The earlier treatment of logarithms and exponential functions did not define the functions precisely and formally. This section develops the concepts in a mathematically rigorous way.
- The cornerstone of the development is the definition of the natural logarithm in terms of an integral.
- The function e^x is then defined as the inverse of the natural logarithm. General exponential functions are defined in terms of e^x, and the corresponding inverse functions are general logarithms.
- Familiar properties of logarithms and exponents still hold in this more rigorous context.

Key Equations

- **Natural logarithm function**
- $\ln x = \int_1^x \frac{1}{t}\, dt$
- **Exponential function** $y = e^x$
- $\ln y = \ln(e^x) = x$

This page titled 2.7: Integrals, Exponential Functions, and Logarithms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.7: Integrals, Exponential Functions, and Logarithms** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.8: Exponential Growth and Decay

Learning Objectives

- Use the exponential growth model in applications, including population growth and compound interest.
- Explain the concept of doubling time.
- Use the exponential decay model in applications, including radioactive decay and Newton's law of cooling.
- Explain the concept of half-life.

One of the most prevalent applications of exponential functions involves growth and decay models. Exponential growth and decay show up in a host of natural applications. From population growth and continuously compounded interest to radioactive decay and Newton's law of cooling, exponential functions are ubiquitous in nature. In this section, we examine exponential growth and decay in the context of some of these applications.

Exponential Growth Model

Many systems exhibit exponential growth. These systems follow a model of the form $y = y_0 e^{kt}$, where y_0 represents the initial state of the system and k is a positive constant, called the growth constant. Notice that in an exponential growth model, we have

$$y' = ky_0 e^{kt} = ky. \tag{2.8.1}$$

That is, the rate of growth is proportional to the current function value. This is a key feature of exponential growth. Equation 2.8.1 involves derivatives and is called *a differential equation.*

Exponential Growth

Systems that exhibit *exponential growth* increase according to the mathematical model

$$y = y_0 e^{kt}$$

where y_0 represents the initial state of the system and $k > 0$ is a constant, called the growth constant.

Population growth is a common example of exponential growth. Consider a population of bacteria, for instance. It seems plausible that the rate of population growth would be proportional to the size of the population. After all, the more bacteria there are to reproduce, the faster the population grows. Figure 2.8.1 and Table 2.8.1 represent the growth of a population of bacteria with an initial population of 200 bacteria and a growth constant of 0.02. Notice that after only 2 hours (120 minutes), the population is 10 times its original size!

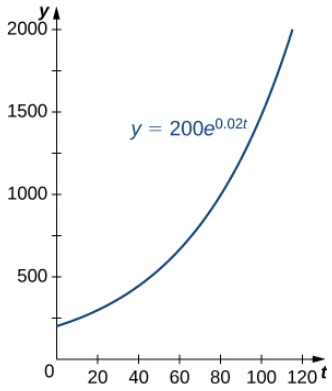


Figure 2.8.1: An example of exponential growth for bacteria.

Table 2.8.1: Exponential Growth of a Bacterial Population

Time(min)	Population Size (no. of bacteria)
10	244
20	298

Time(min)	Population Size (no. of bacteria)
30	364
40	445
50	544
60	664
70	811
80	991
90	1210
100	1478
110	1805
120	2205

Note that we are using a continuous function to model what is inherently discrete behavior. At any given time, the real-world population contains a whole number of bacteria, although the model takes on noninteger values. When using exponential growth models, we must always be careful to interpret the function values in the context of the phenomenon we are modeling.

Example 2.8.1: Population Growth

Consider the population of bacteria described earlier. This population grows according to the function $f(t) = 200e^{0.02t}$, where t is measured in minutes. How many bacteria are present in the population after 5 hours (300 minutes)? When does the population reach $100,000$ bacteria?

Solution

We have $f(t) = 200e^{0.02t}$. Then

$$f(300) = 200e^{0.02(300)} \approx 80,686.$$

There are $80,686$ bacteria in the population after 5 hours.

To find when the population reaches $100,000$ bacteria, we solve the equation

$$\begin{aligned} 100,000 &= 200e^{0.02t} \\ 500 &= e^{0.02t} \\ \ln 500 &= 0.02t \\ t &= \frac{\ln 500}{0.02} \approx 310.73. \end{aligned}$$

The population reaches $100,000$ bacteria after 310.73 minutes.

Exercise 2.8.1

Consider a population of bacteria that grows according to the function $f(t) = 500e^{0.05t}$, where t is measured in minutes. How many bacteria are present in the population after 4 hours? When does the population reach 100 million bacteria?

Answer

Use the process from the previous example.

Answer

There are $81,377,396$ bacteria in the population after 4 hours. The population reaches 100 million bacteria after 244.12 minutes.

Let's now turn our attention to a financial application: **compound interest**. Interest that is not compounded is called simple interest. ***Simple interest*** is paid once, at the end of the specified time period (usually 1 year). So, if we put $\$1000$ in a savings account earning 2 simple interest per year, then at the end of the year we have

$$1000(1+0.02)=\$1020.$$

Compound interest is paid multiple times per year, depending on the compounding period. Therefore, if the bank compounds the interest every 6 months, it credits half of the year's interest to the account after 6 months. During the second half of the year, the account earns interest not only on the initial $\$1000$, but also on the interest earned during the first half of the year. Mathematically speaking, at the end of the year, we have

$$1000\left(1+\frac{0.02}{2}\right)^2=\$1020.10.$$

Similarly, if the interest is compounded every 4 months, we have

$$1000\left(1+\frac{0.02}{3}\right)^3=\$1020.13,$$

and if the interest is compounded daily (365 times per year), we have $\$1020.20$ If we extend this concept, so that the interest is compounded continuously, after t years we have

$$1000\lim_{n\to\infty}\left(1+\frac{0.02}{n}\right)^{nt}.$$

Now let's manipulate this expression so that we have an exponential growth function. Recall that the number e can be expressed as a limit:

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

Based on this, we want the expression inside the parentheses to have the form $(1+1/m)$. Let $n=0.02m$. Note that as $n\to\infty, m\to\infty$ as well. Then we get

$$1000\lim_{n\to\infty}\left(1+\frac{0.02}{n}\right)^{nt}=1000\lim_{m\to\infty}\left(1+\frac{0.02}{0.02m}\right)^{0.02mt}=1000\left[\lim_{m\to\infty}\left(1+\frac{1}{m}\right)^m\right]^{0.02t}.$$

We recognize the limit inside the brackets as the number e. So, the balance in our bank account after t years is given by $1000e^{0.02t}$. Generalizing this concept, we see that if a bank account with an initial balance of $\$P$ earns interest at a rate of r, compounded continuously, then the balance of the account after t years is

$$\text{Balance }=Pe^{rt}.$$

Example 2.8.2: Compound Interest

A 25-year-old student is offered an opportunity to invest some money in a retirement account that pays 5 annual interest compounded continuously. How much does the student need to invest today to have $\$1$ million when she retires at age 65? What if she could earn 6 annual interest compounded continuously instead?

Solution

We have

$$1,000,000=Pe^{0.05(40)}$$

$$P=135,335.28.$$

She must invest $\$135,335.28$at 5 interest.

If, instead, she is able to earn 6 then the equation becomes

$$1,000,000 = Pe^{0.06(40)}$$

$$P = 90,717.95.$$

In this case, she needs to invest only $\$90,717.95$.This is roughly two-thirds the amount she needs to invest at 5. The fact that the interest is compounded continuously greatly magnifies the effect of the 1 increase in interest rate.

Exercise 2.8.2

Suppose instead of investing at age $25\sqrt{b^2-4ac}$, the student waits until age 35. How much would she have to invest at 5? At 6?

Hint

Use the process from the previous example.

Answer

At 5 interest, she must invest $\$223,130.16$At 6 interest, she must invest $\$165,298.89$.

If a quantity grows exponentially, the time it takes for the quantity to double remains constant. In other words, it takes the same amount of time for a population of bacteria to grow from 100 to 200 bacteria as it does to grow from $10,000$ to $20,000$ bacteria. This time is called the doubling time. To calculate the doubling time, we want to know when the quantity reaches twice its original size. So we have

$$\begin{aligned} 2y_0 &= y_0 e^{kt} \\ 2 &= e^{kt} \\ \ln 2 &= kt \\ t &= \frac{\ln 2}{k}. \end{aligned}$$

Definition: Doubling Time

If a quantity grows exponentially, the **doubling time** is the amount of time it takes the quantity to double. It is given by

$$\text{Doubling time} = \frac{\ln 2}{k}.$$

Example 2.8.3: Using the Doubling Time

Assume a population of fish grows exponentially. A pond is stocked initially with 500 fish. After 6 months, there are 1000 fish in the pond. The owner will allow his friends and neighbors to fish on his pond after the fish population reaches $10,000$. When will the owner's friends be allowed to fish?

Solution

We know it takes the population of fish 6 months to double in size. So, if t represents time in months, by the doubling-time formula, we have $6 = (\ln 2)/k$. Then, $k = (\ln 2)/6$. Thus, the population is given by $y = 500e^{((\ln 2)/6)t}$. To figure out when the population reaches $10,000$ fish, we must solve the following equation:

$$\begin{aligned} 10,000 &= 500e^{(\ln 2/6)t} \\ 20 &= e^{(\ln 2/6)t} \\ \ln 20 &= \left(\frac{\ln 2}{6}\right) t \\ t &= \frac{6(\ln 20)}{\ln 2} \\ &\approx 25.93. \end{aligned}$$

The owner's friends have to wait 25.93 months (a little more than 2 years) to fish in the pond.

Exercise 2.8.3

Suppose it takes 9 months for the fish population in Example $2.8.3$ to reach 1000 fish. Under these circumstances, how long do the owner's friends have to wait?

Hint

Use the process from the previous example.

Answer

38.90 months

Exponential Decay Model

Exponential functions can also be used to model populations that shrink (from disease, for example), or chemical compounds that break down over time. We say that such systems exhibit exponential decay, rather than exponential growth. The model is nearly the same, except there is a negative sign in the exponent. Thus, for some positive constant k, we have

$$y = y_0 e^{-kt}.$$

As with exponential growth, there is a differential equation associated with exponential decay. We have

$$y' = -ky_0 e^{-kt} = -ky.$$

Exponential Decay

Systems that exhibit exponential decay behave according to the model

$$y = y_0 e^{-kt},$$

where y_0 represents the initial state of the system and $k > 0$ is a constant, called the decay constant.

Figure $2.8.2$ shows a graph of a representative exponential decay function.

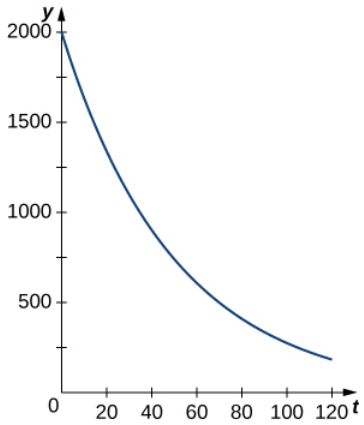


Figure 2.8.2: An example of exponential decay.

Let's look at a physical application of exponential decay. *Newton's law of cooling* says that an object cools at a rate proportional to the difference between the temperature of the object and the temperature of the surroundings. In other words, if T represents the temperature of the object and T_a represents the ambient temperature in a room, then

$$T' = -k(T - T_a).$$

Note that this is not quite the right model for exponential decay. We want the derivative to be proportional to the function, and this expression has the additional T_a term. Fortunately, we can make a change of variables that resolves this issue. Let $y(t) = T(t) - T_a$. Then $y'(t) = T'(t) - 0 = T'(t)$, and our equation becomes

$$y' = -ky.$$

From our previous work, we know this relationship between y and its derivative leads to exponential decay. Thus,

$$y = y_0 e^{-kt},$$

and we see that

$$T - T_a = (T_0 - T_a)e^{-kt}$$

$$T = (T_0 - T_a)e^{-kt} + T_a$$

where T_0 represents the initial temperature. Let's apply this formula in the following example.

Example 2.8.4: Newton's Law of Cooling

According to experienced baristas, the optimal temperature to serve coffee is between $155^\circ F$ and $175^\circ F$. Suppose coffee is poured at a temperature of $200^\circ F$, and after 2 minutes in a $70^\circ F$ room it has cooled to $180^\circ F$. When is the coffee first cool enough to serve? When is the coffee too cold to serve? Round answers to the nearest half minute.

Solution

We have

$$\begin{aligned} T &= (T_0 - T_a)e^{-kt} + T_a \\ 180 &= (200-70)e^{-k(2)} + 70 \\ 110 &= 130e^{-2k} \\ \frac{11}{13} &= e^{-2k} \\ \ln\frac{11}{13} &= -2k \\ \ln 11 - \ln 13 &= -2k \\ k &= \frac{\ln 13 - \ln 11}{2} \end{aligned}$$

Then, the model is

$$T = 130e^{(\ln 11 - \ln 13/2)t} + 70.$$

The coffee reaches $175^\circ F$ when

$$\begin{aligned} 175 &= 130e^{(\ln 11 - \ln 13/2)t} + 70 \\ 105 &= 130e^{(\ln 11 - \ln 13/2)t} \\ \frac{21}{26} &= e^{(\ln 11 - \ln 13/2)t} \\ \ln\frac{21}{26} &= \frac{\ln 11 - \ln 13}{2}t \\ \ln 21 - \ln 26 &= \left(\frac{\ln 11 - \ln 13}{2}\right)t \\ t &= \frac{2(\ln 21 - \ln 26)}{\ln 11 - \ln 13} \\ &\approx 2.56. \end{aligned}$$

The coffee can be served about 2.5 minutes after it is poured. The coffee reaches $155^\circ F$ at

$$\begin{aligned} 155 &= 130e^{(\ln 11 - \ln 13/2)t} + 70 \\ 85 &= 130e^{(\ln 11 - \ln 13)t} \\ \frac{17}{26} &= e^{(\ln 11 - \ln 13)t} \\ \ln 17 - \ln 26 &= \left(\frac{\ln 11 - \ln 13}{2}\right)t \\ t &= \frac{2(\ln 17 - \ln 26)}{\ln 11 - \ln 13} \\ &\approx 5.09. \end{aligned}$$

The coffee is too cold to be served about 5 minutes after it is poured.

? Exercise 2.8.4

Suppose the room is warmer $(75^\circ F)$ and, after 2 minutes, the coffee has cooled only to $185^\circ F$. When is the coffee first cool enough to serve? When is the coffee be too cold to serve? Round answers to the nearest half minute.

Hint

Use the process from the previous example.

Answer

The coffee is first cool enough to serve about 3.5 minutes after it is poured. The coffee is too cold to serve about 7 minutes after it is poured.

Just as systems exhibiting exponential growth have a constant doubling time, systems exhibiting exponential decay have a constant half-life. To calculate the half-life, we want to know when the quantity reaches half its original size. Therefore, we have

$$\begin{aligned}\frac{y_0}{2} &= y_0 e^{-kt}\\ \frac{1}{2} &= e^{-kt}\\ -\ln 2 &= -kt\\ t &= \frac{\ln 2}{k}.\end{aligned}$$

Note: This is the same expression we came up with for doubling time.

Definition: Half-Life

If a quantity decays exponentially, the half-life is the amount of time it takes the quantity to be reduced by half. It is given by

$$\text{Half-life} = \frac{\ln 2}{k}.$$

Example 2.8.5: Radiocarbon Dating

One of the most common applications of an exponential decay model is carbon dating. Carbon-14 decays (emits a radioactive particle) at a regular and consistent exponential rate. Therefore, if we know how much carbon-14 was originally present in an object and how much carbon-14 remains, we can determine the age of the object. The half-life of carbon-14 is approximately 5730 years—meaning, after that many years, half the material has converted from the original carbon-14 to the new nonradioactive nitrogen-14. If we have 100 g carbon-14 today, how much is left in 50 years? If an artifact that originally contained 100 g of carbon-14 now contains 10 g of carbon-14, how old is it? Round the answer to the nearest hundred years.

Solution

We have

$$\begin{aligned}5730 &= \frac{\ln 2}{k}\\ k &= \frac{\ln 2}{5730}.\end{aligned}$$

So, the model says

$$y = 100e^{-(\ln 2/5730)t}.$$

In 50 years, we have

$$y = 100e^{-(\ln 2/5730)(50)} \approx 99.40$$

Therefore, in 50 years, 99.40 g of carbon-14 remains.

To determine the age of the artifact, we must solve

$$\begin{aligned}10 &= 100e^{-(\ln 2/5730)t}\\ \frac{1}{10} &= e^{-(\ln 2/5730)t}\\ t &\approx 19035.\end{aligned}$$

The artifact is about $19,000$ years old.

Exercise 2.8.5: Carbon-14 Decay

If we have 100 g of carbon-14 , how much is left after 500 years? If an artifact that originally contained 100 g of carbon-14 now contains 20 g of carbon-14, how old is it? Round the answer to the nearest hundred years.

Hint

Use the process from the previous example.

Answer

A total of 94.13 g of carbon-14 remains after 500 years. The artifact is approximately 13,300 years old.

Key Concepts

- Exponential growth and exponential decay are two of the most common applications of exponential functions.
- Systems that exhibit exponential growth follow a model of the form $y = y_0 e^{kt}$.
- In exponential growth, the rate of growth is proportional to the quantity present. In other words, $y' = ky$.
- Systems that exhibit exponential growth have a constant doubling time, which is given by $(\ln 2)/k$.
- Systems that exhibit exponential decay follow a model of the form $y = y_0 e^{-kt}$.
- Systems that exhibit exponential decay have a constant half-life, which is given by $(\ln 2)/k$.

Glossary

doubling time

if a quantity grows exponentially, the doubling time is the amount of time it takes the quantity to double, and is given by $(\ln 2)/k$

exponential decay

systems that exhibit exponential decay follow a model of the form $y = y_0 e^{-kt}$

exponential growth

systems that exhibit exponential growth follow a model of the form $y = y_0 e^{kt}$

half-life

if a quantity decays exponentially, the half-life is the amount of time it takes the quantity to be reduced by half. It is given by $(\ln 2)/k$

This page titled 2.8: Exponential Growth and Decay is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.8: Exponential Growth and Decay** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

2.9: Calculus of the Hyperbolic Functions

Learning Objectives

- Apply the formulas for derivatives and integrals of the hyperbolic functions.
- Apply the formulas for the derivatives of the inverse hyperbolic functions and their associated integrals.
- Describe the common applied conditions of a catenary curve.

We were introduced to hyperbolic functions previously, along with some of their basic properties. In this section, we look at differentiation and integration formulas for the hyperbolic functions and their inverses.

Derivatives and Integrals of the Hyperbolic Functions

Recall that the hyperbolic sine and hyperbolic cosine are defined as

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

and

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

The other hyperbolic functions are then defined in terms of $\sinh x$ and $\cosh x$. The graphs of the hyperbolic functions are shown in Figure 2.9.1.

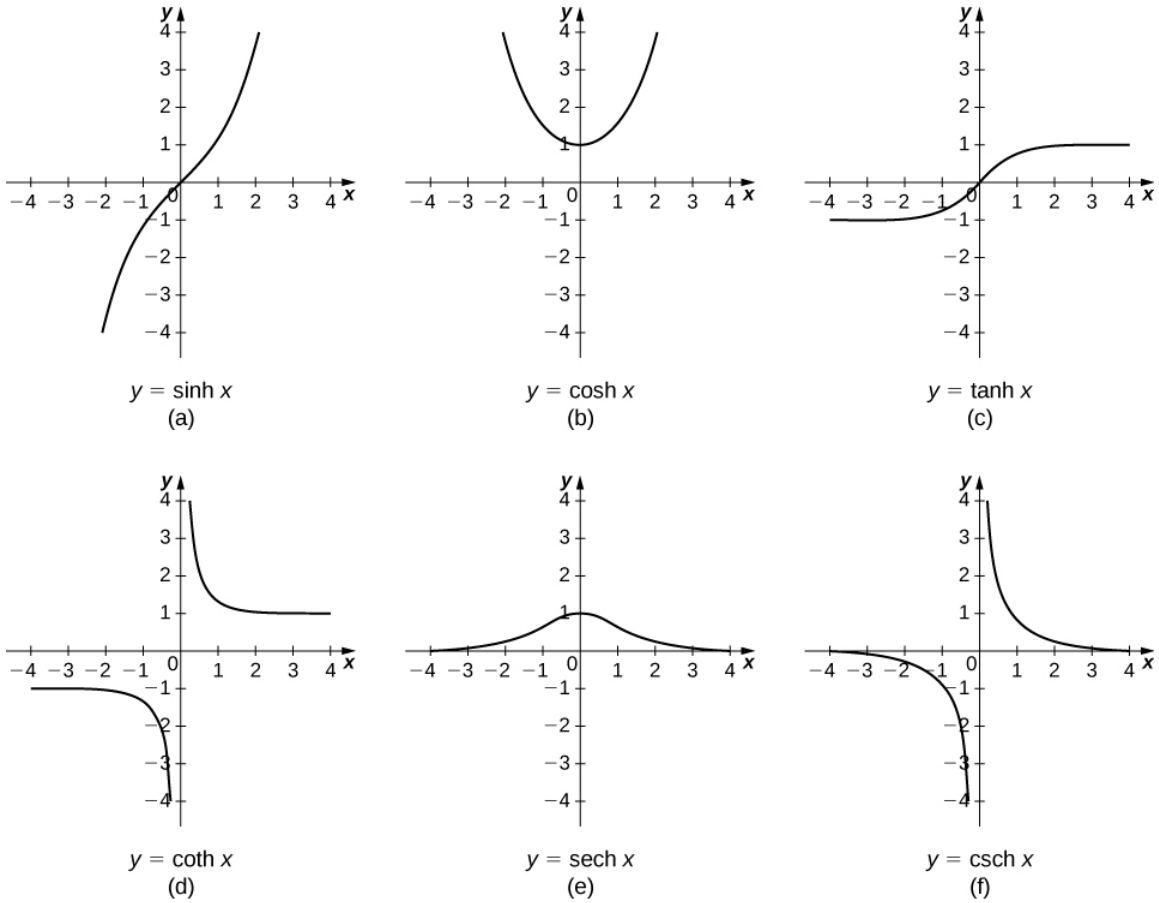


Figure 2.9.1: Graphs of the hyperbolic functions.

It is easy to develop differentiation formulas for the hyperbolic functions. For example, looking at $\sinh x$ we have

$$\begin{aligned}\frac{d}{dx}(\sinh x) &= \frac{d}{dx}\left(\frac{e^x - e^{-x}}{2}\right) \\ &= \frac{1}{2}\left[\frac{d}{dx}(e^x) - \frac{d}{dx}(e^{-x})\right] \\ &= \frac{1}{2}[e^x + e^{-x}] \\ &= \cosh x.\end{aligned}$$

Similarly,

$$\frac{d}{dx}\cosh x = \sinh x.$$

We summarize the differentiation formulas for the hyperbolic functions in Table 2.9.1.

Table 2.9.1: Derivatives of the Hyperbolic Functions

$f(x)$	$\frac{d}{dx}f(x)$
$\sinh x$	$\cosh x$
$\cosh x$	$\sinh x$
$\tanh x$	$\operatorname{sech}^2 x$
$\coth x$	$-\operatorname{csch}^2 x$
$\operatorname{sech} x$	$-\operatorname{sech} x \tanh x$
$\operatorname{csch} x$	$-\operatorname{csch} x \coth x$

Let's take a moment to compare the derivatives of the hyperbolic functions with the derivatives of the standard trigonometric functions. There are a lot of similarities, but differences as well. For example, the derivatives of the sine functions match:

$$\frac{d}{dx}\sin x = \cos x$$

and

$$\frac{d}{dx}\sinh x = \cosh x.$$

The derivatives of the cosine functions, however, differ in sign:

$$\frac{d}{dx}\cos x = -\sin x,$$

but

$$\frac{d}{dx}\cosh x = \sinh x.$$

As we continue our examination of the hyperbolic functions, we must be mindful of their similarities and differences to the standard trigonometric functions. These differentiation formulas for the hyperbolic functions lead directly to the following integral formulas.

$$\int \sinh u\, du = \cosh u + C \tag{2.9.1}$$

$$\int \operatorname{csch}^2 u\, du = -\coth u + C \tag{2.9.2}$$

$$\int \cosh u\, du = \sinh u + C \tag{2.9.3}$$

$$\int \operatorname{sech} u \tanh u\, du = -\operatorname{sech}\ u + C - \operatorname{csch} u + C \tag{2.9.4}$$

$$\int \operatorname{sech}^2 u\, du = \tanh u + C \tag{2.9.5}$$

$$\int \operatorname{csch} u \coth u\, du = -\operatorname{csch} u + C \tag{2.9.6}$$

Example 2.9.1: Differentiating Hyperbolic Functions

Evaluate the following derivatives:

a. $\dfrac{d}{dx}(\sinh(x^2))$

b. $\dfrac{d}{dx}(\cosh x)^2$

Solution:

Using the formulas in Table 2.9.1 and the chain rule, we get

a. $\dfrac{d}{dx}(\sinh(x^2)) = \cosh(x^2) \cdot 2x$

b. $\dfrac{d}{dx}(\cosh x)^2 = 2\cosh x \sinh x$

Exercise 2.9.1

Evaluate the following derivatives:

a. $\dfrac{d}{dx}(\tanh(x^2+3x))$

b. $\dfrac{d}{dx}\left(\dfrac{1}{(\sinh x)^2}\right)$

Hint

Use the formulas in Table 2.9.1and apply the chain rule as necessary.

Answer a

$$\frac{d}{dx}(\tanh(x^2+3x)) = (\operatorname{sech}^2(x^2+3x))(2x+3)$$

Answer b

$$\frac{d}{dx}\left(\frac{1}{(\sinh x)^2}\right) = \frac{d}{dx}(\sinh x)^{-2} = -2(\sinh x)^{-3}\cosh x$$

Example 2.9.2: Integrals Involving Hyperbolic Functions

Evaluate the following integrals:

a. $\displaystyle\int x\cosh(x^2)dx$

b. $\int \tanh x \, dx$

Solution

We can use u-substitution in both cases.

a. Let $u = x^2$. Then, $du = 2x \, dx$ and

$$\begin{aligned} \int x \cosh(x^2) dx &= \int \frac{1}{2} \cosh u \, du \\ &= \frac{1}{2} \sinh u + C \\ &= \frac{1}{2} \sinh(x^2) + C. \end{aligned}$$

b. Let $u = \cosh x$. Then, $du = \sinh x \, dx$ and

$$\begin{aligned} \int \tanh x \, dx &= \int \frac{\sinh x}{\cosh x} \, dx \\ &= \int \frac{1}{u} du \\ &= \ln |u| + C \\ &= \ln |\cosh x| + C. \end{aligned}$$

Note that $\cosh x > 0$ for all x, so we can eliminate the absolute value signs and obtain

$$\int \tanh x \, dx = \ln(\cosh x) + C.$$

? Exercise 2.9.2

Evaluate the following integrals:

a. $\int \sinh^3 x \cosh x \, dx$

b. $\int \text{sech}^2(3x) \, dx$

Hint

Use the formulas above and apply u-substitution as necessary.

Answer a

$$\int \sinh^3 x \cosh x \, dx = \frac{\sinh^4 x}{4} + C$$

Answer b

$$\int \text{sech}^2(3x) \, dx = \frac{\tanh(3x)}{3} + C$$

Calculus of Inverse Hyperbolic Functions

Looking at the graphs of the hyperbolic functions, we see that with appropriate range restrictions, they all have inverses. Most of the necessary range restrictions can be discerned by close examination of the graphs. The domains and ranges of the inverse hyperbolic functions are summarized in Table 2.9.2.

Table 2.9.2: Domains and Ranges of the Inverse Hyperbolic Functions

Function	Domain	Range

Function	Domain	Range
$\sinh^{-1} x$	$(-\infty,\infty)$	$(-\infty,\infty)$
$\cosh^{-1} x$	$(1,\infty)$	$[0,\infty)$
$\tanh^{-1} x$	$(-1,1)$	$(-\infty,\infty)$
$\coth^{-1} x$	$(-\infty,1)\cup(1,\infty)$	$(-\infty,0)\cup(0,\infty)$
$\operatorname{sech}^{-1} x$	$(0,1)$	$[0,\infty)$
$\operatorname{csch}^{-1} x$	$(-\infty,0)\cup(0,\infty)$	$(-\infty,0)\cup(0,\infty)$

The graphs of the inverse hyperbolic functions are shown in the following figure.

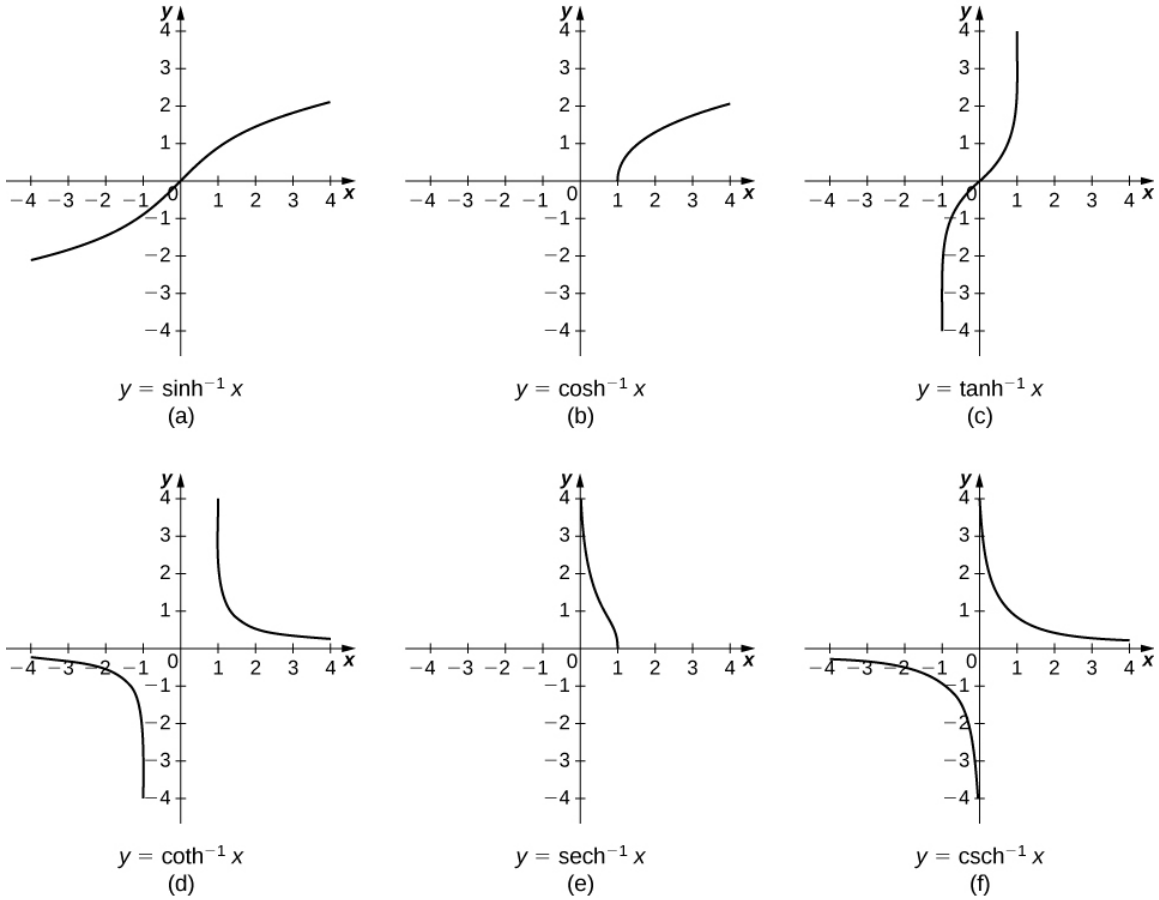


Figure 2.9.3: Graphs of the inverse hyperbolic functions.

To find the derivatives of the inverse functions, we use implicit differentiation. We have

$$y = \sinh^{-1} x \tag{2.9.7}$$

$$\sinh y = x \tag{2.9.8}$$

$$\frac{d}{dx}\sinh y = \frac{d}{dx}x \tag{2.9.9}$$

$$\cosh y \frac{dy}{dx} = 1. \tag{2.9.10}$$

Recall that $\cosh^2 y - \sinh^2 y = 1$, so $\cosh y = \sqrt{1+\sinh^2 y}$. Then,

$$\frac{dy}{dx} = \frac{1}{\cosh y} = \frac{1}{\sqrt{1+\sinh^2 y}} = \frac{1}{\sqrt{1+x^2}}.$$

We can derive differentiation formulas for the other inverse hyperbolic functions in a similar fashion. These differentiation formulas are summarized in Table 2.9.3.

Table 2.9.3: Derivatives of the Inverse Hyperbolic Functions

$f(x)$	$\frac{d}{dx}f(x)$
$\sinh^{-1} x$	$\frac{1}{\sqrt{1+x^2}}$
$\cosh^{-1} x$	$\frac{1}{\sqrt{x^2-1}}$
$\tanh^{-1} x$	$\frac{1}{1-x^2}$
$\coth^{-1} x$	$\frac{1}{1-x^2}$
$\text{sech}^{-1} x$	$\frac{-1}{x\sqrt{1-x^2}}$
$\text{csch}^{-1} x$	$\frac{-1}{\lvert x\rvert\sqrt{1+x^2}}$

Note that the derivatives of $\tanh^{-1} x$ and $\coth^{-1} x$ are the same. Thus, when we integrate $1/(1-x^2)$, we need to select the proper antiderivative based on the domain of the functions and the values of x. Integration formulas involving the inverse hyperbolic functions are summarized as follows.

$$\int \frac{1}{\sqrt{1+u^2}}du = \sinh^{-1} u + C$$

$$\int \frac{1}{u\sqrt{1-u^2}}du = -\text{sech}^{-1}|u| + C$$

$$\int \frac{1}{\sqrt{u^2-1}}du = \cosh^{-1} u + C$$

$$\int \frac{1}{u\sqrt{1+u^2}}du = -\text{csch}^{-1}|u| + C$$

$$\int \frac{1}{1-u^2}du = \begin{cases} \tanh^{-1} u + C & \text{if } |u|<1 \\ \coth^{-1} u + C & \text{if } |u|>1 \end{cases}$$

Example 2.9.3: Differentiating Inverse Hyperbolic Functions

Evaluate the following derivatives:

a. $\frac{d}{dx}\left(\sinh^{-1}\left(\frac{x}{3}\right)\right)$

b. $\frac{d}{dx}\left(\tanh^{-1} x\right)^2$

Solution

Using the formulas in Table 2.9.3 and the chain rule, we obtain the following results:

a. $\frac{d}{dx}(\sinh^{-1}(\frac{x}{3})) = \frac{1}{3\sqrt{1+\frac{x^2}{9}}} = \frac{1}{\sqrt{9+x^2}}$

b. $\dfrac{d}{dx}(\tanh^{-1} x)^2 = \dfrac{2(\tanh^{-1} x)}{1-x^2}$

Exercise 2.9.3

Evaluate the following derivatives:

a. $\dfrac{d}{dx}(\cosh^{-1}(3x))$
b. $\dfrac{d}{dx}(\coth^{-1} x)^3$

Hint

Use the formulas in Table 2.9.3and apply the chain rule as necessary.

Answer a

$$\frac{d}{dx}(\cosh^{-1}(3x)) = \frac{3}{\sqrt{9x^2-1}}$$

Answer b

$$\frac{d}{dx}(\coth^{-1} x)^3 = \frac{3(\coth^{-1} x)^2}{1-x^2}$$

Example 2.9.4: Integrals Involving Inverse Hyperbolic Functions

Evaluate the following integrals:

a. $\displaystyle\int \frac{1}{\sqrt{4x^2-1}}\,dx$
b. $\displaystyle\int \frac{1}{2x\sqrt{1-9x^2}}\,dx$

Solution

We can use ***u*-substitution** in both cases.

Let $u = 2x$. Then, $du = 2\,dx$ and we have

$$\begin{aligned}\int \frac{1}{\sqrt{4x^2-1}}\,dx &= \int \frac{1}{2\sqrt{u^2-1}}\,du \\ &= \frac{1}{2}\cosh^{-1} u + C \\ &= \frac{1}{2}\cosh^{-1}(2x) + C.\end{aligned}$$

Let $u = 3x$. Then, $du = 3\,dx$ and we obtain

$$\begin{aligned}\int \frac{1}{2x\sqrt{1-9x^2}}\,dx &= \frac{1}{2}\int \frac{1}{u\sqrt{1-u^2}}\,du \\ &= -\frac{1}{2}\operatorname{sech}^{-1}|u| + C \\ &= -\frac{1}{2}\operatorname{sech}^{-1}|3x| + C\end{aligned}$$

https://math.libretexts.org/@go/page/33795

Exercise 2.9.4

Evaluate the following integrals:

a. $\int \frac{1}{\sqrt{x^2-4}}dx, x>2$

b. $\int \frac{1}{\sqrt{1-e^{2x}}}dx$

Hint

Use the formulas above and apply u-substitution as necessary.

Answer a

$$\int \frac{1}{\sqrt{x^2-4}}dx = \cosh^{-1}(\frac{x}{2})+C$$

Answer b

$$\int \frac{1}{\sqrt{1-e^{2x}}}dx = -\text{sech}^{-1}(e^x)+C$$

Applications

One physical application of hyperbolic functions involves hanging cables. If a cable of uniform density is suspended between two supports without any load other than its own weight, the cable forms a curve called a **catenary**. High-voltage power lines, chains hanging between two posts, and strands of a spider's web all form catenaries. The following figure shows chains hanging from a row of posts.



Figure 2.9.3: Chains between these posts take the shape of a catenary. (credit: modification of work by OKFoundryCompany, Flickr)

Hyperbolic functions can be used to model catenaries. Specifically, functions of the form $y = a \cdot \cosh(x/a)$ are catenaries. Figure 2.9.4 shows the graph of $y = 2\cosh(x/2)$.

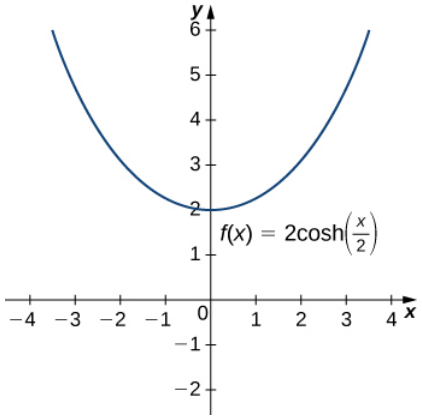


Figure 2.9.4: A hyperbolic cosine function forms the shape of a catenary.

Example 2.9.5: Using a Catenary to Find the Length of a Cable

Assume a hanging cable has the shape $10\cosh(x/10)$ for $-15 \le x \le 15$, where x is measured in feet. Determine the length of the cable (in feet).

Solution

Recall from Section 6.4 that the formula for arc length is

$$\underbrace{\int_a^b \sqrt{1+[f'(x)]^2}\,dx.}_{\text{Arc Length}}$$

We have $f(x) = 10\cosh(x/10)$, so $f'(x) = \sinh(x/10)$. Then the arc length is

$$\int_a^b \sqrt{1+[f'(x)]^2}\,dx = \int_{-15}^{15} \sqrt{1+\sinh^2\left(\frac{x}{10}\right)}\,dx.$$

Now recall that

$$1+\sinh^2 x = \cosh^2 x,$$

so we have

$$\begin{aligned}
\text{Arc Length} &= \int_{-15}^{15} \sqrt{1+\sinh^2\left(\frac{x}{10}\right)}\,dx \\
&= \int_{-15}^{15} \cosh\left(\frac{x}{10}\right)dx \\
&= 10\sinh\left(\frac{x}{10}\right)\Big|_{-15}^{15} \\
&= 10\left[\sinh\left(\frac{3}{2}\right) - \sinh\left(-\frac{3}{2}\right)\right] \\
&= 20\sinh\left(\frac{3}{2}\right) \\
&\approx 42.586 \text{ ft.}
\end{aligned}$$

Exercise 2.9.5:

Assume a hanging cable has the shape $15\cosh(x/15)$ for $-20 \le x \le 20$. Determine the length of the cable (in feet).

Answer

52.95ft

Key Concepts

- Hyperbolic functions are defined in terms of exponential functions.
- Term-by-term differentiation yields differentiation formulas for the hyperbolic functions. These differentiation formulas give rise, in turn, to integration formulas.
- With appropriate range restrictions, the hyperbolic functions all have inverses.
- Implicit differentiation yields differentiation formulas for the inverse hyperbolic functions, which in turn give rise to integration formulas.
- The most common physical applications of hyperbolic functions are calculations involving catenaries.

Glossary

catenary

a curve in the shape of the function $y = a \cdot \cosh(x/a)$ is a catenary; a cable of uniform density suspended between two supports assumes the shape of a catenary

---

This page titled 2.9: Calculus of the Hyperbolic Functions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **6.9: Calculus of the Hyperbolic Functions** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

CHAPTER OVERVIEW

3: Techniques of Integration

We saw in the previous chapter how important integration can be for all kinds of different topics—from calculations of volumes to flow rates, and from using a velocity function to determine a position to locating centers of mass. It is no surprise, then, that techniques for finding antiderivatives (or indefinite integrals) are important to know for everyone who uses them. We have already discussed some basic integration formulas and the method of integration by substitution. In this chapter, we study some additional techniques, including some ways of approximating definite integrals when normal techniques do not work.

3.E: Techniques of Integration (Exercises)
3.0: Prelude to Techniques of Integration
3.1: Integration by Parts
3.2: Trigonometric Integrals
3.3: Trigonometric Substitution
3.4: Partial Fractions
3.5: Other Strategies for Integration
3.6: Numerical Integration
3.7: Improper Integrals

Contributors

Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 3: Techniques of Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

3.E: Techniques of Integration (Exercises)

This page titled 3.E: Techniques of Integration (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

3.0: Prelude to Techniques of Integration

In a large city, accidents occurred at an average rate of one every three months at a particularly busy intersection. After residents complained, changes were made to the traffic lights at the intersection. It has now been eight months since the changes were made and there have been no accidents. Were the changes effective or is the eight-month interval without an accident a result of chance? We explore this question later in this chapter and see that integration is an essential part of determining the answer.



Figure 3.0.1: Careful planning of traffic signals can prevent or reduce the number of accidents at busy intersections. (credit: modification of work by David McKelvey, Flickr)

We saw in the previous chapter how important integration can be for all kinds of different topics—from calculations of volumes to flow rates, and from using a velocity function to determine a position to locating centers of mass. It is no surprise, then, that techniques for finding antiderivatives (or indefinite integrals) are important to know for everyone who uses them. We have already discussed some basic integration formulas and the method of integration by substitution. In this chapter, we study some additional techniques, including some ways of approximating definite integrals when normal techniques do not work.

Template:HideTOC

This page titled 3.0: Prelude to Techniques of Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.0: Prelude to Techniques of Integration** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.1: Integration by Parts

Learning Objectives

- Recognize when to use integration by parts.
- Use the integration-by-parts formula to solve integration problems.
- Use the integration-by-parts formula for definite integrals.

By now we have a fairly thorough procedure for how to evaluate many basic integrals. However, although we can integrate $\int x\sin(x^2)\,dx$ by using the substitution, $u=x^2$, something as simple looking as $\int x\sin x\;dx$ defies us. Many students want to know whether there is a product rule for integration. There is not, but there is a technique based on the product rule for differentiation that allows us to exchange one integral for another. We call this technique **integration by parts.**

The Integration-by-Parts Formula

If, $h(x)=f(x)g(x)$, then by using the product rule, we obtain

$$h'(x)=f'(x)g(x)+g'(x)f(x). \tag{3.1.1}$$

Although at first it may seem counterproductive, let's now integrate both sides of Equation 3.1.1:

$$\int h'(x)\;dx=\int (g(x)f'(x)+f(x)g'(x))\;dx.$$

This gives us

$$h(x)=f(x)g(x)=\int g(x)f'(x)\,dx+\int f(x)g'(x)\;dx.$$

Now we solve for $\int f(x)g'(x)\;dx$:

$$\int f(x)g'(x)\,dx=f(x)g(x)-\int g(x)f'(x)\;dx.$$

By making the substitutions $u=f(x)$ and $v=g(x)$, which in turn make $du=f'(x)\,dx$ and $dv=g'(x)\,dx$, we have the more compact form

$$\int u\,dv=uv-\int v\,du.$$

Integration by Parts

Let $u=f(x)$ and $v=g(x)$ be functions with continuous derivatives. Then, the integration-by-parts formula for the integral involving these two functions is:

$$\int u\,dv=uv-\int v\,du. \tag{3.1.2}$$

The advantage of using the integration-by-parts formula is that we can use it to exchange one integral for another, possibly easier, integral. The following example illustrates its use.

Example 3.1.1: Using Integration by Parts

Use integration by parts with $u=x$ and $dv=\sin x\;dx$ to evaluate

$$\int x\sin x\;dx.$$

Solution

By choosing $u=x$, we have $du=1\;dx$. Since $dv=\sin x\;dx$, we get

$$v = \int \sin x \; dx = -\cos x.$$

It is handy to keep track of these values as follows:

- $u = x$
- $dv = \sin x \; dx$
- $du = 1 \, dx$
- $v = \int \sin x \; dx = -\cos x.$

Applying the integration-by-parts formula (Equation 3.1.2) results in

$$\begin{aligned} \int x \sin x \; dx &= (x)(-\cos x) - \int (-\cos x)(1 \; dx) && \text{(Substitute)} \\ &= -x \cos x + \int \cos x \; dx && \text{(Simplify)} \end{aligned}$$

Then use

$$\int \cos x \; dx = \sin x + C.$$

to obtain

$$\int x \sin x \; dx = -x \cos x + \sin x + C.$$

Analysis

At this point, there are probably a few items that need clarification. First of all, you may be curious about what would have happened if we had chosen $u = \sin x$ and $dv = x \, dx$. If we had done so, then we would have $du = \cos x$ and $v = \frac{1}{2}x^2$. Thus, after applying integration by parts (Equation 3.1.2), we have

$$\int x \sin x \; dx = \frac{1}{2}x^2 \sin x - \int \frac{1}{2}x^2 \cos x \; dx.$$

Unfortunately, with the new integral, we are in no better position than before. It is important to keep in mind that when we apply integration by parts, we may need to try several choices for u and dv before finding a choice that works.

Second, you may wonder why, when we find $v = \int \sin x \; dx = -\cos x$, we do not use $v = -\cos x + K$. To see that it makes no difference, we can rework the problem using $v = -\cos x + K$:

$$\begin{aligned} \int x \sin x \; dx &= (x)(-\cos x + K) - \int (-\cos x + K)(1 \; dx) \\ &= -x \cos x + Kx + \int \cos x \; dx - \int K \; dx \\ &= -x \cos x + Kx + \sin x - Kx + C \\ &= -x \cos x + \sin x + C. \end{aligned}$$

As you can see, it makes no difference in the final solution.

Last, we can check to make sure that our antiderivative is correct by differentiating $-x \cos x + \sin x + C$:

$$\begin{aligned} \frac{d}{dx}(-x \cos x + \sin x + C) &= \cancel{(-1)\cos x} + (-x)(-\sin x) + \cancel{\cos x} \\ &= x \sin x \end{aligned}$$

Therefore, the antiderivative checks out.

? Exercise 3.1.1

Evaluate $\int xe^{2x}\,dx$ using the integration-by-parts formula (Equation 3.1.2) with $u = x$ and $dv = e^{2x}\,dx$.

Hint

Find du and v, and use the previous example as a guide.

Answer

$$\int xe^{2x}\,dx = \frac{1}{2}xe^{2x} - \frac{1}{4}e^{2x} + C$$

The natural question to ask at this point is: How do we know how to choose u and dv? Sometimes it is a matter of trial and error; however, the acronym **LIATE** can often help to take some of the guesswork out of our choices. This acronym stands for **L**ogarithmic Functions, **I**nverse Trigonometric Functions, **A**lgebraic Functions, **T**rigonometric Functions, and **E**xponential Functions. This mnemonic serves as an aid in determining an appropriate choice for u. The type of function in the integral that appears first in the list should be our first choice of u.

For example, if an integral contains a **logarithmic function** and an **algebraic function,** we should choose u to be the logarithmic function, because L comes before A in LIATE. The integral in Example 3.1.1 has a trigonometric function ($\sin x$) and an algebraic function (x). Because A comes before T in LIATE, we chose u to be the algebraic function. When we have chosen u, dv is selected to be the remaining part of the function to be integrated, together with dx.

Why does this mnemonic work? Remember that whatever we pick to be dv must be something we can integrate. Since we do not have integration formulas that allow us to integrate simple logarithmic functions and inverse trigonometric functions, it makes sense that they should not be chosen as values for dv. Consequently, they should be at the head of the list as choices for u. Thus, we put LI at the beginning of the mnemonic. (We could just as easily have started with IL, since these two types of functions won't appear together in an integration-by-parts problem.) The exponential and trigonometric functions are at the end of our list because they are fairly easy to integrate and make good choices for dv. Thus, we have TE at the end of our mnemonic. (We could just as easily have used ET at the end, since when these types of functions appear together it usually doesn't really matter which one is u and which one is dv.) Algebraic functions are generally easy both to integrate and to differentiate, and they come in the middle of the mnemonic.

✓ Example 3.1.2: Using Integration by Parts

Evaluate

$$\int \frac{\ln x}{x^3}\,dx.$$

Solution

Begin by rewriting the integral:

$$\int \frac{\ln x}{x^3}\,dx = \int x^{-3}\ln x\,dx.$$

Since this integral contains the algebraic function x^{-3} and the logarithmic function $\ln x$, choose $u = \ln x$, since L comes before A in LIATE. After we have chosen $u = \ln x$, we must choose $dv = x^{-3}\,dx$.

Next, since $u = \ln x$, we have $du = \frac{1}{x}\,dx$. Also, $v = \int x^{-3}\,dx = -\frac{1}{2}x^{-2}$. Summarizing,

- $u = \ln x$
- $du = \frac{1}{x}\,dx$
- $dv = x^{-3}\,dx$
- $v = \int x^{-3}\,dx = -\frac{1}{2}x^{-2}$.

Substituting into the integration-by-parts formula (Equation 3.1.2) gives

$$\begin{aligned}\int \frac{\ln x}{x^3}\,dx &= \int x^{-3}\ln x\,dx = (\ln x)(-\frac{1}{2}x^{-2}) - \int (-\frac{1}{2}x^{-2})(\frac{1}{x}\,dx)\\ &= -\frac{1}{2}x^{-2}\ln x + \int \frac{1}{2}x^{-3}\,dx\\ &= -\frac{1}{2}x^{-2}\ln x - \frac{1}{4}x^{-2} + C\\ &= -\frac{1}{2x^2}\ln x - \frac{1}{4x^2} + C\end{aligned}$$

? Exercise 3.1.2

Evaluate

$$\int x\ln x\,\,dx.$$

Hint

Use $u = \ln x$ and $dv = x\,\,dx$.

Answer

$$\int x\ln x\,\,dx = \frac{1}{2}x^2\ln x - \frac{1}{4}x^2 + C$$

In some cases, as in the next two examples, it may be necessary to apply integration by parts more than once.

✓ Example 3.1.3A: Applying Integration by Parts More Than Once

Evaluate

$$\int x^2 e^{3x}\,dx.$$

Solution

Using LIATE, choose $u = x^2$ and $dv = e^{3x}\,dx$. Thus, $du = 2x\,dx$ and $v = \int e^{3x}\,dx = \left(\frac{1}{3}\right)e^{3x}$. Therefore,

- $u = x^2$
- $du = 2x\,dx$
- $dv = e^{3x}\,dx$
- $v = \int e^{3x}\,dx = \frac{1}{3}e^{3x}$.

Substituting into Equation 3.1.2 produces

$$\int x^2 e^{3x}\,dx = \frac{1}{3}x^2 e^{3x} - \int \frac{2}{3}xe^{3x}\,dx. \tag{3.1.3}$$

We still cannot integrate $\int \frac{2}{3}xe^{3x}\,dx$ directly, but the integral now has a lower power on x. We can evaluate this new integral by using integration by parts again. To do this, choose

$$u = x$$

and

$$dv = \frac{2}{3}e^{3x}\,dx.$$

Thus,

$$du = dx$$

and

$$v = \int \left(\frac{2}{3}\right) e^{3x}\, dx = \left(\frac{2}{9}\right) e^{3x}.$$

Now we have

- $u = x$
- $du = dx$
- $dv = \frac{2}{3} e^{3x}\, dx$
- $v = \int \frac{2}{3} e^{3x}\, dx = \frac{2}{9} e^{3x}.$

Substituting back into Equation 3.1.3 yields

$$\int x^2 e^{3x}\, dx = \frac{1}{3} x^2 e^{3x} - \left(\frac{2}{9} x e^{3x} - \int \frac{2}{9} e^{3x}\, dx\right).$$

After evaluating the last integral and simplifying, we obtain

$$\int x^2 e^{3x}\, dx = \frac{1}{3} x^2 e^{3x} - \frac{2}{9} x e^{3x} + \frac{2}{27} e^{3x} + C.$$

Example 3.1.3B: Applying Integration by Parts When LIATE Does not Quite Work

Evaluate

$$\int t^3 e^{t^2}\, dt.$$

Solution

If we use a strict interpretation of the mnemonic LIATE to make our choice of u, we end up with $u = t^3$ and $dv = e^{t^2}\, dt$. Unfortunately, this choice won't work because we are unable to evaluate $\int e^{t^2}\, dt$. However, since we can evaluate $\int t e^{t^2}\, dx$, we can try choosing $u = t^2$ and $dv = t e^{t^2}\, dt$. With these choices we have

- $u = t^2$
- $du = 2t dt$
- $dv = t e^{t^2}\, dt$
- $v = \int t e^{t^2}\, dt = \frac{1}{2} e^{t^2}.$

Thus, we obtain

$$\begin{aligned}\int t^3 e^{t^2}\, dt &= \frac{1}{2} t^2 e^{t^2} - \int \frac{1}{2} e^{t^2} 2t\, dt \\ &= \frac{1}{2} t^2 e^{t^2} - \frac{1}{2} e^{t^2} + C.\end{aligned}$$

Example 3.1.3C: Applying Integration by Parts More Than Once

Evaluate

$$\int \sin(\ln x)\, dx.$$

Solution

This integral appears to have only one function—namely, $\sin(\ln x)$—however, we can always use the constant function 1 as the other function. In this example, let's choose $u = \sin(\ln x)$ and $dv = 1\, dx$. (The decision to use $u = \sin(\ln x)$ is easy. We

can't choose $dv = \sin(\ln x)\,dx$ because if we could integrate it, we wouldn't be using integration by parts in the first place!) Consequently, $du = (1/x)\cos(\ln x)\,dx$ and $v = \int 1\,dx = x$. After applying integration by parts to the integral and simplifying, we have

$$\int \sin(\ln x)\,dx = x\sin(\ln x) - \int \cos(\ln x)\,dx.$$

Unfortunately, this process leaves us with a new integral that is very similar to the original. However, let's see what happens when we apply integration by parts again. This time let's choose $u = \cos(\ln x)$ and $dv = 1\,dx$, making $du = -(1/x)\sin(\ln x)\,dx$ and $v = \int 1\,dx = x$.

Substituting, we have

$$\int \sin(\ln x)\,dx = x\sin(\ln x) - (x\cos(\ln x) - \int -\sin(\ln x)\,dx).$$

After simplifying, we obtain

$$\int \sin(\ln x)\,dx = x\sin(\ln x) - x\cos(\ln x) - \int \sin(\ln x)\,dx.$$

The last integral is now the same as the original. It may seem that we have simply gone in a circle, but now we can actually evaluate the integral. To see how to do this more clearly, substitute $I = \int \sin(\ln x)\,dx$. Thus, the equation becomes

$$I = x\sin(\ln x) - x\cos(\ln x) - I.$$

First, add I to both sides of the equation to obtain

$$2I = x\sin(\ln x) - x\cos(\ln x).$$

Next, divide by 2:

$$I = \frac{1}{2}x\sin(\ln x) - \frac{1}{2}x\cos(\ln x).$$

Substituting $I = \int \sin(\ln x)\,dx$ again, we have

$$\int \sin(\ln x)\,dx = \frac{1}{2}x\sin(\ln x) - \frac{1}{2}x\cos(\ln x).$$

From this we see that $(1/2)x\sin(\ln x) - (1/2)x\cos(\ln x)$ is an antiderivative of $\sin(\ln x)\,dx$. For the most general antiderivative, add $+C$:

$$\int \sin(\ln x)\,dx = \frac{1}{2}x\sin(\ln x) - \frac{1}{2}x\cos(\ln x) + C.$$

Analysis

If this method feels a little strange at first, we can check the answer by differentiation:

$$\begin{aligned}\frac{d}{dx}\left(\frac{1}{2}x\sin(\ln x) - \frac{1}{2}x\cos(\ln x)\right) &= \frac{1}{2}(\sin(\ln x)) + \cos(\ln x)\cdot\frac{1}{x}\cdot\frac{1}{2}x - \left(\frac{1}{2}\cos(\ln x) - \sin(\ln x)\cdot\frac{1}{x}\cdot\frac{1}{2}x\right)\\ &= \sin(\ln x).\end{aligned}$$

Exercise 3.1.3

Evaluate

$$\int x^2 \sin x\,dx.$$

Hint

This is similar to Examples $3.1.3A$- $3.1.3C$.

Answer

$$\int x^2 \sin x \, dx = -x^2 \cos x + 2x \sin x + 2 \cos x + C$$

Integration by Parts for Definite Integrals

Now that we have used integration by parts successfully to evaluate *indefinite integrals*, we turn our attention to definite integrals. The integration technique is really the same, only we add a step to evaluate the integral at the upper and lower limits of integration.

Integration by Parts for Definite Integrals

Let $u = f(x)$ and $v = g(x)$ be functions with continuous derivatives on $[a, b]$. Then

$$\int_a^b u \, dv = uv\Big|_a^b - \int_a^b v \, du$$

Example $3.1.4A$: Finding the Area of a Region

Find the area of the region bounded above by the graph of $y = \tan^{-1} x$ and below by the x-axis over the interval $[0, 1]$.

Solution

This region is shown in Figure $3.1.1$. To find the area, we must evaluate

$$\int_0^1 \tan^{-1} x \; dx.$$

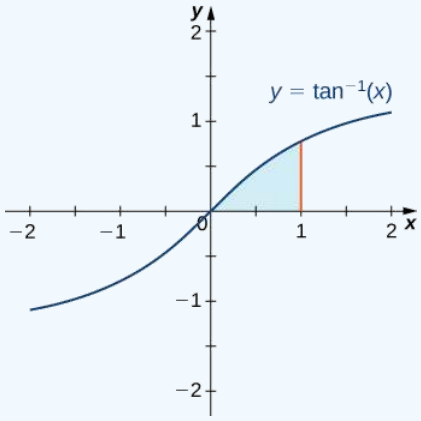


Figure 3.1.1: To find the area of the shaded region, we have to use integration by parts.

For this integral, let's choose $u = tan^{-1} x$ and $dv = dx$, thereby making $du = \dfrac{1}{x^2+1} dx$ and $v = x$. After applying the integration-by-parts formula (Equation 3.1.2) we obtain

$$\text{Area} = x \tan^{-1} x\Big|_0^1 - \int_0^1 \frac{x}{x^2+1} \, dx.$$

Use u-substitution to obtain

$$\int_0^1 \frac{x}{x^2+1} \, dx = \frac{1}{2} \ln\left(x^2+1\right)\Big|_0^1.$$

Thus,

$$\text{Area} = x\tan^{-1}x\Big|_0^1 - \frac{1}{2}\ln(x^2+1)\Big|_0^1 = \left(\frac{\pi}{4} - \frac{1}{2}\ln 2\right) \text{ units}^2.$$

At this point it might not be a bad idea to do a "reality check" on the reasonableness of our solution. Since $\frac{\pi}{4} - \frac{1}{2}\ln 2 \approx 0.4388 \text{ units}^2$, and from Figure 3.1.1 we expect our area to be slightly less than 0.5 units^2, this solution appears to be reasonable.

Example 3.1.4B: Finding a Volume of Revolution

Find the volume of the solid obtained by revolving the region bounded by the graph of $f(x) = e^{-x}$, the x-axis, the y-axis, and the line $x = 1$ about the y-axis.

Solution

The best option to solving this problem is to use the shell method. Begin by sketching the region to be revolved, along with a typical rectangle (Figure 3.1.2).

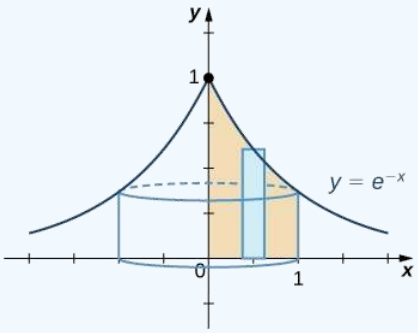


Figure 3.1.2: We can use the shell method to find a volume of revolution.

To find the volume using shells, we must evaluate

$$2\pi \int_0^1 xe^{-x}\,dx. \tag{3.1.4}$$

To do this, let $u = x$ and $dv = e^{-x}$. These choices lead to $du = dx$ and $v = \int e^{-x}\,dx = -e^{-x}$. Using the Shell Method formula, we obtain

$$\begin{aligned}
\text{Volume} &= 2\pi \int_0^1 xe^{-x}\,dx \\
&= 2\pi\left(-xe^{-x}\Big|_0^1 + \int_0^1 e^{-x}\,dx\right) && \text{(Use integration by parts)} \\
&= 2\pi\left(-e^{-1} + 0 - e^{-x}\Big|_0^1\right) \\
&= 2\pi\left(-e^{-1} - e^{-1} + 1\right) \\
&= 2\pi\left(1 - \frac{2}{e}\right) \text{ units}^3. && \text{(Evaluate and simplify)}
\end{aligned}$$

Analysis

Again, it is a good idea to check the reasonableness of our solution. We observe that the solid has a volume slightly less than that of a cylinder of radius 1 and height of $1/e$ added to the volume of a cone of base radius 1 and height of $1 - \frac{1}{e}$. Consequently, the solid should have a volume a bit less than

$$\pi(1)^2\frac{1}{e} + \left(\frac{\pi}{3}\right)(1)^2\left(1 - \frac{1}{e}\right) = \frac{2\pi}{3e} + \frac{\pi}{3} \approx 1.8177 \text{ units}^3.$$

Since $2\pi - \frac{4\pi}{e} \approx 1.6603$, we see that our calculated volume is reasonable.

? Exercise 3.1.4

Evaluate

$$\int_0^{\pi/2} x \cos x \, dx.$$

Hint

Use Equation 3.1.2with $u = x$ and $dv = \cos x \, dx$.

Answer

$$\int_0^{\pi/2} x \cos x \, dx = \frac{\pi}{2} - 1$$

Key Concepts

- The integration-by-parts formula (Equation 3.1.2) allows the exchange of one integral for another, possibly easier, integral.
- Integration by parts applies to both definite and indefinite integrals.

Key Equations

- **Integration by parts formula**

  $$\int u \, dv = uv - \int v \, du$$

- **Integration by parts for definite integrals**

  $$\int_a^b u \, dv = uv\Big|_a^b - \int_a^b v \, du$$

Glossary

integration by parts

a technique of integration that allows the exchange of one integral for another using the formula $\int u \, dv = uv - \int v \, du$

This page titled 3.1: Integration by Parts is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.1: Integration by Parts** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.2: Trigonometric Integrals

Learning Objectives

- Solve integration problems involving products and powers of $\sin x$ and $\cos x$.
- Solve integration problems involving products and powers of $\tan x$ and $\sec x$.
- Use reduction formulas to solve trigonometric integrals.

In this section we look at how to integrate a variety of products of trigonometric functions. These integrals are called **trigonometric integrals**. They are an important part of the integration technique called *trigonometric substitution*, which is featured in Trigonometric Substitution. This technique allows us to convert algebraic expressions that we may not be able to integrate into expressions involving trigonometric functions, which we may be able to integrate using the techniques described in this section. In addition, these types of integrals appear frequently when we study polar, cylindrical, and spherical coordinate systems later. Let's begin our study with products of $\sin x$ and $\cos x$.

Integrating Products and Powers of sin x and cos x

A key idea behind the strategy used to integrate combinations of products and powers of $\sin x$ and $\cos x$ involves rewriting these expressions as sums and differences of integrals of the form $\int \sin^j x \cos x \, dx$ or $\int \cos^j x \sin x \, dx$. After rewriting these integrals, we evaluate them using u-substitution. Before describing the general process in detail, let's take a look at the following examples.

Example 3.2.1: Integrating $\int \cos^j x \sin x \, dx$

Evaluate $\int \cos^3 x \sin x \, dx$.

Solution

Use u-substitution and let $u = \cos x$. In this case, $du = -\sin x \, dx$.

Thus,

$$\int \cos^3 x \sin x \, dx = -\int u^3 \, du = -\frac{1}{4}u^4 + C = -\frac{1}{4}\cos^4 x + C.$$

Exercise 3.2.1

Evaluate $\int \sin^4 x \cos x \, dx$.

Hint

Let $u = \sin x$.

Answer

$$\int \sin^4 x \cos x \, dx = \frac{1}{5}\sin^5 x + C$$

Example 3.2.2: A Preliminary Example: Integrating $\int \cos^j x \sin^k x \, dx$ where k is Odd

Evaluate $\int \cos^2 x \sin^3 x \, dx$.

Solution

To convert this integral to integrals of the form $\int \cos^j x \sin x \, dx$, rewrite $\sin^3 x = \sin^2 x \sin x$ and make the substitution $\sin^2 x = 1 - \cos^2 x$.

Thus,

$$\int \cos^2 x \sin^3 x\, dx = \int \cos^2 x (1-\cos^2 x) \sin x\, dx \qquad \text{Let } u = \cos x;\ \text{then } du = -\sin x\, dx.$$

$$= -\int u^2 (1-u^2)\, du$$

$$= \int (u^4 - u^2)\, du$$

$$= \frac{1}{5}u^5 - \frac{1}{3}u^3 + C$$

$$= \frac{1}{5}\cos^5 x - \frac{1}{3}\cos^3 x + C.$$

? Exercise 3.2.2

Evaluate $\int \cos^3 x \sin^2 x\, dx$.

Hint

Write $\cos^3 x = \cos^2 x \cos x = (1-\sin^2 x)\cos x$ and let $u = \sin x$.

Answer

$$\int \cos^3 x \sin^2 x\, dx = \frac{1}{3}\sin^3 x - \frac{1}{5}\sin^5 x + C$$

In the next example, we see the strategy that must be applied when there are only even powers of $\sin x$ and $\cos x$. For integrals of this type, the identities

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

and

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

are invaluable. These identities are sometimes known as **power-reducing identities** and they may be derived from the double-angle identity $\cos(2x) = \cos^2 x - \sin^2 x$ and the Pythagorean identity $\cos^2 x + \sin^2 x = 1$.

✓ Example 3.2.3: Integrating an Even Power of $\sin x$

Evaluate $\int \sin^2 x\, dx$.

Solution

To evaluate this integral, let's use the trigonometric identity $\sin^2 x = \frac{1}{2} - \frac{1}{2}\cos(2x)$. Thus,

$$\int \sin^2 x\, dx = \int \left(\frac{1}{2} - \frac{1}{2}\cos(2x)\right) dx = \frac{1}{2}x - \frac{1}{4}\sin(2x) + C.$$

? Exercise 3.2.3

Evaluate $\int \cos^2 x\, dx$.

Hint

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

Answer

$$\int \cos^2 x\, dx = \frac{1}{2}x + \frac{1}{4}\sin(2x) + C$$

The general process for integrating products of powers of $\sin x$ and $\cos x$ is summarized in the following set of guidelines.

Problem-Solving Strategy: Integrating Products and Powers of $\sin x$ and $cosx$

To integrate $\int \cos^j x \sin^k x\, dx$ use the following strategies:

1. If k is odd, rewrite $\sin^k x = \sin^{k-1} x \sin x$ and use the identity $\sin^2 x = 1 - \cos^2 x$ to rewrite $\sin^{k-1} x$ in terms of $\cos x$. Integrate using the substitution $u = \cos x$. This substitution makes $du = -\sin x\, dx$.

2. If j is odd, rewrite $\cos^j x = \cos^{j-1} x \cos x$ and use the identity $\cos^2 x = 1 - \sin^2 x$ to rewrite $\cos^{j-1} x$ in terms of $\sin x$. Integrate using the substitution $u = \sin x$. This substitution makes $du = \cos x\, dx$. (Note: If both j and k are odd, either strategy 1 or strategy 2 may be used.)

3. If both j and k are even, use $\sin^2 x = \dfrac{1 - \cos(2x)}{2}$ and $\cos^2 x = \dfrac{1 + \cos(2x)}{2}$. After applying these formulas, simplify and reapply strategies 1 through 3 as appropriate.

Example 3.2.4: Integrating $\int \cos^j x \sin^k x\, dx$ where k is Odd

Evaluate $\int \cos^8 x \sin^5 x\, dx$.

Solution

Since the power on $\sin x$ is odd, use strategy 1. Thus,

$$\begin{aligned}
\int \cos^8 x \sin^5 x\, dx &= \int \cos^8 x \sin^4 x \sin x\, dx && \text{Break off } \sin x. \\
&= \int \cos^8 x (\sin^2 x)^2 \sin x\, dx && \text{Rewrite } \sin^4 x = (\sin^2 x)^2. \\
&= \int \cos^8 x (1 - \cos^2 x)^2 \sin x\, dx && \text{Substitute } \sin^2 x = 1 - \cos^2 x. \\
&= \int u^8 (1 - u^2)^2 (-du) && \text{Let } u = \cos x \text{ and } du = -\sin x\, dx. \\
&= \int (-u^8 + 2u^{10} - u^{12}) du && \text{Expand.} \\
&= -\frac{1}{9}u^9 + \frac{2}{11}u^{11} - \frac{1}{13}u^{13} + C && \text{Evaluate the integral.} \\
&= -\frac{1}{9}\cos^9 x + \frac{2}{11}\cos^{11} x - \frac{1}{13}\cos^{13} x + C && \text{Substitute } u = \cos x.
\end{aligned}$$

Example 3.2.5: Integrating $\int \cos^j x \sin^k x\, dx$ where k and j are Even

Evaluate $\int \sin^4 x\, dx$.

Solution: Since the power on $\sin x$ is even $(k = 4)$ and the power on $\cos x$ is even $(j = 0)$, we must use strategy 3. Thus,

$$\int \sin^4 x\,dx = \int \left(\sin^2 x\right)^2 dx \qquad \text{Rewrite } \sin^4 x = \left(\sin^2 x\right)^2.$$

$$= \int \left(\frac{1}{2} - \frac{1}{2}\cos(2x)\right)^2 dx \qquad \text{Substitute } \sin^2 x = \frac{1}{2} - \frac{1}{2}\cos(2x).$$

$$= \int \left(\frac{1}{4} - \frac{1}{2}\cos(2x) + \frac{1}{4}\cos^2(2x)\right) dx \qquad \text{Expand } \left(\frac{1}{2} - \frac{1}{2}\cos(2x)\right)^2.$$

$$= \int \left(\frac{1}{4} - \frac{1}{2}\cos(2x) + \frac{1}{4}\left(\frac{1}{2} + \frac{1}{2}\cos(4x)\right)\right) dx \qquad \text{Use Power Reduction Formula on } \cos^2(2x).$$

$$= \int \left(\frac{3}{8} - \frac{1}{2}\cos(2x) + \frac{1}{8}\cos(4x)\right) dx \qquad \text{Simplify.}$$

$$= \frac{3}{8}x - \frac{1}{4}\sin(2x) + \frac{1}{32}\sin(4x) + C \qquad \text{Evaluate the integral.}$$

? Exercise 3.2.4

Evaluate $\int \cos^3 x\,dx$.

Hint

Use strategy 2. Write $\cos^3 x = \cos^2 x \cos x$ and substitute $\cos^2 x = 1 - \sin^2 x$.

Answer

$$\int \cos^3 x\,dx = \sin x - \frac{1}{3}\sin^3 x + C$$

? Exercise 3.2.5

Evaluate $\int \cos^2(3x)\,dx$.

Hint

Use strategy 3. Substitute $\cos^2(3x) = \frac{1}{2} + \frac{1}{2}\cos(6x)$

Answer

$$\int \cos^2(3x)\,dx = \frac{1}{2}x + \frac{1}{12}\sin(6x) + C$$

Product to Sum Formulas

In some areas of physics, such as quantum mechanics, signal processing, and the computation of Fourier series, it is often necessary to integrate products that include $sin(ax), sin(bx), cos(ax)$, and $cos(bx)$. These integrals are evaluated by applying trigonometric identities, as outlined in the following rule.

Rule: Integrating Products of Sines and Cosines of Different Angles

To integrate products involving $\sin(ax)$, $\sin(bx)$, $\cos(ax)$, and $\cos(bx)$, use the substitutions

$$\sin(ax)\sin(bx) = \frac{1}{2}\cos((a-b)x) - \frac{1}{2}\cos((a+b)x)$$

$$\sin(ax)\cos(bx) = \frac{1}{2}\sin((a-b)x) + \frac{1}{2}\sin((a+b)x)$$

$$\cos(ax)\cos(bx) = \frac{1}{2}\cos((a-b)x) + \frac{1}{2}\cos((a+b)x)$$

These formulas may be derived from the sum-of-angle formulas for sine and cosine.

Example 3.2.6: Evaluating $\int \sin(ax)\cos(bx)\,dx$

Evaluate $\int \sin(5x)\cos(3x)\,dx$.

Solution: Apply the identity $\sin(5x)\cos(3x) = \frac{1}{2}\sin(2x) + \frac{1}{2}\sin(8x)$. Thus,

$$\int \sin(5x)\cos(3x)\,dx = \int \frac{1}{2}\sin(2x) + \frac{1}{2}\sin(8x)\,dx = -\frac{1}{4}\cos(2x) - \frac{1}{16}\cos(8x) + C.$$

Exercise 3.2.6

Evaluate $\int \cos(6x)\cos(5x)\,dx$.

Hint

Substitute $\cos(6x)\cos(5x) = \frac{1}{2}\cos x + \frac{1}{2}\cos(11x)$.

Answer

$$\int \cos(6x)\cos(5x)\,dx = \frac{1}{2}\sin x + \frac{1}{22}\sin(11x) + C$$

Integrating Products and Powers of tan x and sec x

Before discussing the integration of products and powers of $\tan x$ and $\sec x$, it is useful to recall the integrals involving $\tan x$ and $\sec x$ we have already learned:

1. $\int \sec^2 x\,dx = \tan x + C$
2. $\int \sec x \tan x\,dx = \sec x + C$
3. $\int \tan x\,dx = \ln|\sec x| + C$
4. $\int \sec x\,dx = \ln|\sec x + \tan x| + C.$

For most integrals of products and powers of $\tan x$ and $\sec x$, we rewrite the expression we wish to integrate as the sum or difference of integrals of the form $\int \tan^j x \sec^2 x\,dx$ or $\int \sec^j x \tan x\,dx$. As we see in the following example, we can evaluate these new integrals by using u-substitution.

Example 3.2.7: Evaluating $\int \sec^j x \tan x\,dx$

Evaluate $\int \sec^5 x \tan x\,dx$.

Solution: Start by rewriting $\sec^5 x \tan x$ as $\sec^4 x \sec x \tan x$.

$$\begin{aligned}\int \sec^5 x \tan x\, dx &= \int \sec^4 x \sec x \tan x\, dx \\ &= \int u^4\, du && \text{Let } u = \sec x;\ \text{then, } du = \sec x \tan x\, dx. \\ &= \tfrac{1}{5}u^5 + C && \text{Evaluate the integral.} \\ &= \tfrac{1}{5}\sec^5 x + C && \text{Substitute } \sec x = u.\end{aligned}$$

You can read some interesting information at this website to learn about a common integral involving the secant.

? Exercise 3.2.7

Evaluate $\int \tan^5 x \sec^2 x\, dx$.

Hint

Let $u = \tan x$ and $du = \sec^2 x$.

Answer

$$\int \tan^5 x \sec^2 x\, dx = \tfrac{1}{6}\tan^6 x + C$$

We now take a look at the various strategies for integrating products and powers of $\sec x$ and $\tan x$.

Problem-Solving Strategy: Integrating $\int \tan^k x \sec^j x\, dx$

To integrate $\int \tan^k x \sec^j x\, dx$, use the following strategies:

1. If j is even and $j \geq 2$, rewrite $\sec^j x = \sec^{j-2} x \sec^2 x$ and use $\sec^2 x = \tan^2 x + 1$ to rewrite $\sec^{j-2} x$ in terms of $\tan x$. Let $u = \tan x$ and $du = \sec^2 x$.

2. If k is odd and $j \geq 1$, rewrite $\tan^k x \sec^j x = \tan^{k-1} x \sec^{j-1} x \sec x \tan x$ and use $\tan^2 x = \sec^2 x - 1$ to rewrite $\tan^{k-1} x$ in terms of $\sec x$. Let $u = \sec x$ and $du = \sec x \tan x\, dx$. (Note: If j is even and k is odd, then either strategy 1 or strategy 2 may be used.)

3. If k is odd where $k \geq 3$ and $j = 0$, rewrite $\tan^k x = \tan^{k-2} x \tan^2 x = \tan^{k-2} x(\sec^2 x - 1) = \tan^{k-2} x \sec^2 x - \tan^{k-2} x$. It may be necessary to repeat this process on the $\tan^{k-2} x$ term.

4. If k is even and j is odd, then use $\tan^2 x = \sec^2 x - 1$ to express $\tan^k x$ in terms of $\sec x$. Use integration by parts to integrate odd powers of $\sec x$.

✓ Example 3.2.8: Integrating $\int \tan^k x \sec^j x\, dx$ when j is Even

Evaluate $\int \tan^6 x \sec^4 x\, dx$.

Solution

Since the power on $\sec x$ is even, rewrite $\sec^4 x = \sec^2 x \sec^2 x$ and use $\sec^2 x = \tan^2 x + 1$ to rewrite the first $\sec^2 x$ in terms of $\tan x$. Thus,

$$\begin{aligned}
\int \tan^6 x \sec^4 x\, dx &= \int \tan^6 x(\tan^2 x+1)\sec^2 x\, dx \\
&= \int u^6(u^2+1)\, du && \text{Let } u=\tan x \text{ and } du=\sec^2 x. \\
&= \int (u^8+u^6)\, du && \text{Expand.} \\
&= \frac{1}{9}u^9+\frac{1}{7}u^7+C && \text{Evaluate the integral.} \\
&= \frac{1}{9}\tan^9 x+\frac{1}{7}\tan^7 x+C. && \text{Substitute } \tan x=u.
\end{aligned}$$

Example 3.2.9: Integrating $\int \tan^k x \sec^j x\, dx$ when k is Odd

Evaluate $\int \tan^5 x \sec^3 x\, dx$.

Solution

Since the power on $\tan x$ is odd, begin by rewriting $\tan^5 x \sec^3 x = \tan^4 x \sec^2 x \sec x \tan x$. Thus,

$$\begin{aligned}
\int \tan^5 x \sec^3 x\, dx &= \int \tan^4 x \sec^2 x \sec x \tan x\, dx. \\
&= \int (\tan^2 x)^2 \sec^2 x \sec x \tan x\, dx && \text{Write } \tan^4 x=(\tan^2 x)^2. \\
&= \int (\sec^2 x-1)^2 \sec^2 x \sec x \tan x\, dx && \text{Use } \tan^2 x=\sec^2 x-1. \\
&= \int (u^2-1)^2 u^2\, du && \text{Let } u=\sec x \text{ and } du=\sec x \tan x\, dx \\
&= \int (u^6-2u^4+u^2)du && \text{Expand.} \\
&= \frac{1}{7}u^7-\frac{2}{5}u^5+\frac{1}{3}u^3+C && \text{Integrate.} \\
&= \frac{1}{7}\sec^7 x-\frac{2}{5}\sec^5 x+\frac{1}{3}\sec^3 x+C && \text{Substitute } \sec x=u.
\end{aligned}$$

Example 3.2.10: Integrating $\int \tan^k x\, dx$ where k is Odd and $k \geq 3$

Evaluate $\int \tan^3 x\, dx$.

Solution

Begin by rewriting $\tan^3 x = \tan x \tan^2 x = \tan x(\sec^2 x-1) = \tan x \sec^2 x - \tan x$. Thus,

$$\begin{aligned}
\int \tan^3 x\, dx &= \int (\tan x \sec^2 x - \tan x)\, dx \\
&= \int \tan x \sec^2 x\, dx - \int \tan x\, dx \\
&= \frac{1}{2}\tan^2 x - \ln|\sec x| + C.
\end{aligned}$$

For the first integral, use the substitution $u=\tan x$. For the second integral, use the formula.

Example 3.2.11: Integrating $\int \sec^3 x\,dx$

Integrate $\int \sec^3 x\,dx$.

Solution

This integral requires integration by parts. To begin, let $u = \sec x$ and $dv = \sec^2 x$. These choices make $du = \sec x \tan x$ and $v = \tan x$. Thus,

$$\begin{aligned}
\int \sec^3 x\,dx &= \sec x \tan x - \int \tan x \sec x \tan x\,dx \\
&= \sec x \tan x - \int \tan^2 x \sec x\,dx && \text{Simplify.} \\
&= \sec x \tan x - \int (\sec^2 x - 1)\sec x\,dx && \text{Substitute } \tan^2 x = \sec^2 x - 1. \\
&= \sec x \tan x + \int \sec x\,dx - \int \sec^3 x\,dx && \text{Rewrite.} \\
&= \sec x \tan x + \ln|\sec x + \tan x| - \int \sec^3 x\,dx. && \text{Evaluate } \int \sec x\,dx.
\end{aligned}$$

We now have

$$\int \sec^3 x\,dx = \sec x \tan x + \ln|\sec x + \tan x| - \int \sec^3 x\,dx.$$

Since the integral $\int \sec^3 x\,dx$ has reappeared on the right-hand side, we can solve for $\int \sec^3 x\,dx$ by adding it to both sides. In doing so, we obtain

$$2\int \sec^3 x\,dx = \sec x \tan x + \ln|\sec x + \tan x|.$$

Dividing by 2, we arrive at

$$\int \sec^3 x\,dx = \frac{1}{2}\sec x \tan x + \frac{1}{2}\ln|\sec x + \tan x| + C$$

Exercise 3.2.8

Evaluate $\int \tan^3 x \sec^7 x\,dx$.

Hint

Use Example 3.2.9as a guide.

Answer

$$\int \tan^3 x \sec^7 x\,dx = \frac{1}{9}\sec^9 x - \frac{1}{7}\sec^7 x + C$$

Reduction Formulas

Evaluating $\int \sec^n x\,dx$ for values of n where n is odd requires integration by parts. In addition, we must also know the value of $\int \sec^{n-2} x\,dx$ to evaluate $\int \sec^n x\,dx$. The evaluation of $\int \tan^n x\,dx$ also requires being able to integrate $\int \tan^{n-2} x\,dx$. To

make the process easier, we can derive and apply the following power reduction formulas. These rules allow us to replace the integral of a power of $\sec x$ or $\tan x$ with the integral of a lower power of $\sec x$ or $\tan x$.

Rule: Reduction Formulas for $\int \sec^n x\,dx$ and $\int \tan^n x\,dx$

$$\int \sec^n x\,dx = \frac{1}{n-1}\sec^{n-2} x \tan x + \frac{n-2}{n-1}\int \sec^{n-2} x\,dx$$

$$\int \tan^n x\,dx = \frac{1}{n-1}\tan^{n-1} x - \int \tan^{n-2} x\,dx$$

The first power reduction rule may be verified by applying integration by parts. The second may be verified by following the strategy outlined for integrating odd powers of $\tan x$.

Example 3.2.12: Revisiting $\int \sec^3 x\,dx$

Apply a reduction formula to evaluate $\int \sec^3 x\,dx$.

Solution: By applying the first reduction formula, we obtain

$$\begin{aligned}\int \sec^3 x\,dx &= \frac{1}{2}\sec x \tan x + \frac{1}{2}\int \sec x\,dx \\ &= \frac{1}{2}\sec x \tan x + \frac{1}{2}\ln|\sec x + \tan x| + C.\end{aligned}$$

Example 3.2.13: Using a Reduction Formula

Evaluate $\int \tan^4 x\,dx$.

Solution: Applying the reduction formula for $\int \tan^4 x\,dx$ we have

$$\begin{aligned}\int \tan^4 x\,dx &= \frac{1}{3}\tan^3 x - \int \tan^2 x\,dx \\ &= \frac{1}{3}\tan^3 x - (\tan x - \int \tan^0 x\,dx) && \text{Apply the reduction formula to } \int \tan^2 x\,dx. \\ &= \frac{1}{3}\tan^3 x - \tan x + \int 1\,dx && \text{Simplify.} \\ &= \frac{1}{3}\tan^3 x - \tan x + x + C && \text{Evaluate } \int 1\,dx\end{aligned}$$

Exercise 3.2.9

Apply the reduction formula to $\int \sec^5 x\,dx$.

Hint

Use reduction formula 1 and let $n = 5$.

Answer

$$\int \sec^5 x\,dx = \frac{1}{4}\sec^3 x \tan x + \frac{3}{4}\int \sec^3 x$$

Key Concepts

Integrals of trigonometric functions can be evaluated by the use of various strategies. These strategies include

1. Applying trigonometric identities to rewrite the integral so that it may be evaluated by u-substitution
2. Using integration by parts
3. Applying trigonometric identities to rewrite products of sines and cosines with different arguments as the sum of individual sine and cosine functions
4. Applying reduction formulas

Key Equations

To integrate products involving $\sin(ax)$, $\sin(bx)$, $\cos(ax)$,and $\cos(bx)$, use the substitutions.

- **Sine Products**

  $$\sin(ax)\sin(bx) = \tfrac{1}{2}\cos((a-b)x) - \tfrac{1}{2}\cos((a+b)x)$$

- **Sine and Cosine Products**

  $$\sin(ax)\cos(bx) = \tfrac{1}{2}\sin((a-b)x) + \tfrac{1}{2}\sin((a+b)x)$$

- **Cosine Products**

  $$\cos(ax)\cos(bx) = \tfrac{1}{2}\cos((a-b)x) + \tfrac{1}{2}\cos((a+b)x)$$

- **Power Reduction Formula**

  $$\int \sec^n x\,dx = \frac{1}{n-1}\sec^{n-2} x\tan x + \frac{n-2}{n-1}\int \sec^{n-2} x\,dx$$

- **Power Reduction Formula**

  $$\int \tan^n x\,dx = \frac{1}{n-1}\tan^{n-1} x - \int \tan^{n-2} x\,dx$$

Glossary

power reduction formula
a rule that allows an integral of a power of a trigonometric function to be exchanged for an integral involving a lower power

trigonometric integral
an integral involving powers and products of trigonometric functions

---

This page titled 3.2: Trigonometric Integrals is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.2: Trigonometric Integrals** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.3: Trigonometric Substitution

Learning Objectives

- Solve integration problems involving the square root of a sum or difference of two squares.

In this section, we explore integrals containing expressions of the form $\sqrt{a^2-x^2}$, $\sqrt{a^2+x^2}$, and $\sqrt{x^2-a^2}$, where the values of a are positive. We have already encountered and evaluated integrals containing some expressions of this type, but many still remain inaccessible. The technique of *trigonometric substitution* comes in very handy when evaluating these integrals. This technique uses substitution to rewrite these integrals as trigonometric integrals.

Integrals Involving $\sqrt{a^2-x^2}$

Before developing a general strategy for integrals containing $\sqrt{a^2-x^2}$, consider the integral $\int \sqrt{9-x^2}\,dx$. This integral cannot be evaluated using any of the techniques we have discussed so far. However, if we make the substitution $x=3\sin\theta$, we have $dx=3\cos\theta\,d\theta$. After substituting into the integral, we have

$$\int \sqrt{9-x^2}\,dx=\int \sqrt{9-(3\sin\theta)^2}\cdot 3\cos\theta\,d\theta.$$

After simplifying, we have

$$\int \sqrt{9-x^2}\,dx=\int 9\sqrt{1-\sin^2\theta}\cdot\cos\theta\,d\theta.$$

Letting $1-\sin^2\theta=\cos^2\theta$, we now have

$$\int \sqrt{9-x^2}\,dx=\int 9\sqrt{\cos^2\theta}\cos\theta\,d\theta.$$

Assuming that $\cos\theta\geq 0$, we have

$$\int \sqrt{9-x^2}\,dx=\int 9\cos^2\theta\,d\theta.$$

At this point, we can evaluate the integral using the techniques developed for integrating powers and products of trigonometric functions. Before completing this example, let's take a look at the general theory behind this idea.

To evaluate integrals involving $\sqrt{a^2-x^2}$, we make the substitution $x=a\sin\theta$ and $dx=a\cos\theta$. To see that this actually makes sense, consider the following argument: The domain of $\sqrt{a^2-x^2}$ is $[-a,a]$. Thus,

$$-a\leq x\leq a.$$

Consequently,

$$-1\leq\frac{x}{a}\leq 1.$$

Since the range of $\sin x$ over $[-(\pi/2),\pi/2]$is $[-1,1]$, there is a unique angle θ satisfying $-(\pi/2)\leq\theta\leq\pi/2$ so that $\sin\theta=x/a$, or equivalently, so that $x=a\sin\theta$. If we substitute $x=a\sin\theta$ into $\sqrt{a^2-x^2}$, we get

$$\begin{aligned}
\sqrt{a^2-x^2} &= \sqrt{a^2-(a\sin\theta)^2} && \text{Let } x = a\sin\theta \text{ where } -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}. \\
& && \text{Simplify.} \\
&= \sqrt{a^2-a^2\sin^2\theta} && \text{Factor out } a^2. \\
&= \sqrt{a^2(1-\sin^2\theta)} && \text{Substitute } 1-\sin^2 x = \cos^2 x. \\
&= \sqrt{a^2\cos^2\theta} && \text{Take the square root.} \\
&= |a\cos\theta| \\
&= a\cos\theta
\end{aligned}$$

Since $\cos x \ge 0$ on $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$ and $a > 0$, $|a\cos\theta| = a\cos\theta$. We can see, from this discussion, that by making the substitution $x = a\sin\theta$, we are able to convert an integral involving a radical into an integral involving trigonometric functions. After we evaluate the integral, we can convert the solution back to an expression involving x. To see how to do this, let's begin by assuming that $0 < x < a$. In this case, $0 < \theta < \frac{\pi}{2}$. Since $\sin\theta = \frac{x}{a}$, we can draw the reference triangle in Figure 3.3.1 to assist in expressing the values of $\cos\theta$, $\tan\theta$, and the remaining trigonometric functions in terms of x. It can be shown that this triangle actually produces the correct values of the trigonometric functions evaluated at θ for all θ satisfying $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$. It is useful to observe that the expression $\sqrt{a^2-x^2}$ actually appears as the length of one side of the triangle. Last, should θ appear by itself, we use $\theta = \sin^{-1}\left(\frac{x}{a}\right)$.

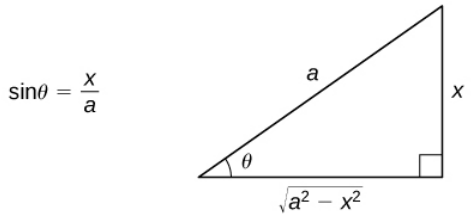


Figure 3.3.1: A reference triangle can help express the trigonometric functions evaluated at θ in terms of x.

The essential part of this discussion is summarized in the following problem-solving strategy.

Problem-Solving Strategy: Integrating Expressions Involving $\sqrt{a^2-x^2}$

1. It is a good idea to make sure the integral cannot be evaluated easily in another way. For example, although this method can be applied to integrals of the form $\int \frac{1}{\sqrt{a^2-x^2}}\,dx$, $\int \frac{x}{\sqrt{a^2-x^2}}\,dx$, and $\int x\sqrt{a^2-x^2}\,dx$, they can each be integrated directly either by formula or by a simple u-substitution.
2. Make the substitution $x = a\sin\theta$ and $dx = a\cos\theta\,d\theta$. Note: This substitution yields $\sqrt{a^2-x^2} = a\cos\theta$.
3. Simplify the expression.
4. Evaluate the integral using techniques from the section on trigonometric integrals.
5. Use the reference triangle from Figure 1 to rewrite the result in terms of x. You may also need to use some trigonometric identities and the relationship $\theta = \sin^{-1}\left(\frac{x}{a}\right)$.

The following example demonstrates the application of this problem-solving strategy.

Example 3.3.1: Integrating an Expression Involving $\sqrt{a^2-x^2}$

Evaluate

$$\int \sqrt{9-x^2}\,dx.$$

Solution

Begin by making the substitutions $x = 3\sin\theta$ and $dx = 3\cos\theta\, d\theta$. Since $\sin\theta = \dfrac{x}{3}$, we can construct the reference triangle shown in Figure 2.

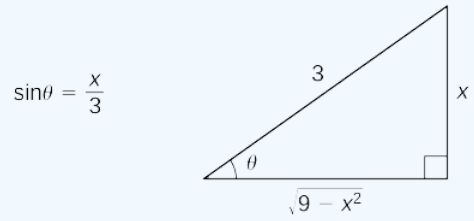


Figure 3.3.2: A reference triangle can be constructed for Example 3.3.1.

Thus,

$$\begin{aligned}
\int \sqrt{9-x^2}\, dx &= \int \sqrt{9-(3\sin\theta)^2}\, 3\cos\theta\, d\theta \quad \text{Substitute } x = 3\sin\theta \text{ and } dx = 3\cos\theta\, d\theta. \\
&= \int \sqrt{9(1-\sin^2\theta)} \cdot 3\cos\theta\, d\theta \quad \text{Simplify.} \\
&= \int \sqrt{9\cos^2\theta} \cdot 3\cos\theta\, d\theta \quad \text{Substitute } \cos^2\theta = 1-\sin^2\theta\,. \\
&= \int 3|\cos\theta| 3\cos\theta\, d\theta \quad \text{Take the square root.} \\
&= \int 9\cos^2\theta\, d\theta \quad \text{Simplify. Since } -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, \cos\theta \ge 0 \text{ and } |\cos\theta| = \cos\theta. \\
&= \int 9\left(\frac{1}{2} + \frac{1}{2}\cos(2\theta)\right) d\theta \quad \text{Use the strategy for integrating an even power of } \cos\theta. \\
&= \frac{9}{2}\theta + \frac{9}{4}\sin(2\theta) + C \quad \text{Evaluate the integral.} \\
&= \frac{9}{2}\theta + \frac{9}{4}(2\sin\theta\cos\theta) + C \quad \text{Substitute } \sin(2\theta) = 2\sin\theta\cos\theta\,. \\
&= \frac{9}{2}\sin^{-1}\left(\frac{x}{3}\right) + \frac{9}{2}\cdot\frac{x}{3}\cdot\frac{\sqrt{9-x^2}}{3} + C \quad \text{Substitute } \sin^{-1}\left(\frac{x}{3}\right) = \theta \text{ and } \sin\theta = \tfrac{x}{3}\,. \\
&\qquad \text{Use the reference triangle to see that } \cos\theta = \frac{\sqrt{9-x^2}}{3} \text{ and make this substitution. Simplify.} \\
&= \frac{9}{2}\sin^{-1}\left(\frac{x}{3}\right) + \frac{x\sqrt{9-x^2}}{2} + C. \quad \text{Simplify.}
\end{aligned}$$

Example 3.3.2: Integrating an Expression Involving $\sqrt{a^2 - x^2}$

Evaluate

$$\int \frac{\sqrt{4-x^2}}{x}\, dx.$$

Solution

First make the substitutions $x = 2\sin\theta$ and $dx = 2\cos\theta\, d\theta$. Since $\sin\theta = \dfrac{x}{2}$, we can construct the reference triangle shown in Figure 3.3.3.

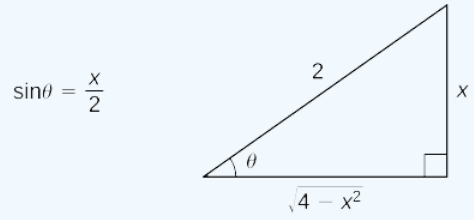


Figure 3.3.3: A reference triangle can be constructed for Example 3.3.2.

Thus,

$$\int \frac{\sqrt{4-x^2}}{x}\,dx = \int \frac{\sqrt{4-(2\sin\theta)^2}}{2\sin\theta}\,2\cos\theta\,d\theta \text{ Substitute } x=2\sin\theta \text{ and } dx=2\cos\theta\,d\theta.$$

$$= \int \frac{2\cos^2\theta}{\sin\theta}\,d\theta \text{ Substitute } \cos^2\theta = 1-\sin^2\theta \text{ and simplify.}$$

$$= \int \frac{2(1-\sin^2\theta)}{\sin\theta}\,d\theta \text{ Substitute } \cos^2\theta = 1-\sin^2\theta\,.$$

$$= \int (2\csc\theta - 2\sin\theta)\,d\theta \text{ Separate the numerator, simplify, and use } \csc\theta = \frac{1}{\sin\theta}\,.$$

$$= 2\ln|\csc\theta - \cot\theta| + 2\cos\theta + C \text{ Evaluate the integral.}$$

$$= 2\ln\left|\frac{2}{x} - \frac{\sqrt{4-x^2}}{x}\right| + \sqrt{4-x^2} + C. \text{ Use the reference triangle to rewrite the expression in terms of } x \text{ and simplify.}$$

In the next example, we see that we sometimes have a choice of methods.

Example 3.3.3: Integrating an Expression Involving $\sqrt{a^2-x^2}$ Two Ways

Evaluate $\int x^3\sqrt{1-x^2}\,dx$ two ways: first by using the substitution $u=1-x^2$ and then by using a trigonometric substitution.

Method 1

Let $u=1-x^2$ and hence $x^2=1-u$. Thus, $du=-2x\,dx$. In this case, the integral becomes

$$\int x^3\sqrt{1-x^2}\,dx = -\frac{1}{2}\int x^2\sqrt{1-x^2}\,(-2x\,dx) \text{ Make the substitution.}$$

$$= -\frac{1}{2}\int (1-u)\sqrt{u}\,du \text{ Expand the expression.}$$

$$= -\frac{1}{2}\int (u^{1/2}-u^{3/2})\,du \text{ Evaluate the integral.}$$

$$= -\frac{1}{2}\left(\frac{2}{3}u^{3/2} - \frac{2}{5}u^{5/2}\right) + C \text{ Rewrite in terms of x.}$$

$$= -\frac{1}{3}(1-x^2)^{3/2} + \frac{1}{5}(1-x^2)^{5/2} + C.$$

Method 2

Let $x=\sin\theta$. In this case, $dx=\cos\theta\,d\theta$. Using this substitution, we have

$$\int x^3\sqrt{1-x^2}\,dx = \int \sin^3\theta\cos^2\theta\,d\theta$$

$$= \int (1-\cos^2\theta)\cos^2\theta\sin\theta\,d\theta \text{ Let } u=\cos\theta. \text{ Thus,} du=-\sin\theta\,d\theta.$$

$$= \int (u^4-u^2)\,du$$

$$= \frac{1}{5}u^5 - \frac{1}{3}u^3 + C \text{ Substitute } \cos\theta = u.$$

$$= \frac{1}{5}\cos^5\theta - \frac{1}{3}\cos^3\theta + C \text{ Use a reference triangle to see that } \cos\theta = \sqrt{1-x^2}\,.$$

$$= \frac{1}{5}(1-x^2)^{5/2} - \frac{1}{3}(1-x^2)^{3/2} + C.$$

Exercise 3.3.1

Rewrite the integral $\displaystyle\int \frac{x^3}{\sqrt{25-x^2}}\,dx$ using the appropriate trigonometric substitution (do not evaluate the integral).

Hint

Substitute $x = 5\sin\theta$ and $dx = 5\cos\theta\, d\theta$.

Answer

$$\int 125\sin^3\theta\, d\theta$$

Integrating Expressions Involving $\sqrt{a^2 + x^2}$

For integrals containing $\sqrt{a^2+x^2}$,let's first consider the domain of this expression. Since $\sqrt{a^2+x^2}$ is defined for all real values of x, we restrict our choice to those trigonometric functions that have a range of all real numbers. Thus, our choice is restricted to selecting either $x = a\tan\theta$ or $x = a\cot\theta$. Either of these substitutions would actually work, but the standard substitution is $x = a\tan\theta$ or, equivalently, $\tan\theta = x/a$. With this substitution, we make the assumption that $-(\pi/2) < \theta < \pi/2$, so that we also have $\theta = \tan^{-1}(x/a)$. The procedure for using this substitution is outlined in the following problem-solving strategy.

Problem-Solving Strategy: Integrating Expressions Involving $\sqrt{a^2 + x^2}$

1. Check to see whether the integral can be evaluated easily by using another method. In some cases, it is more convenient to use an alternative method.
2. Substitute $x = a\tan\theta$ and $dx = a\sec^2\theta\, d\theta$. This substitution yields $\sqrt{a^2+x^2} = \sqrt{a^2+(a\tan\theta)^2} = \sqrt{a^2(1+\tan^2\theta)} = \sqrt{a^2 sec^2\theta} = |a\sec\theta| = a\sec\theta$. (Since $-\dfrac{\pi}{2} < \theta < \dfrac{\pi}{2}$ and $\sec\theta > 0$ over this interval, $|a\sec\theta| = a\sec\theta$.)
3. Simplify the expression.
4. Evaluate the integral using techniques from the section on trigonometric integrals.
5. Use the reference triangle from Figure 3.3.4 to rewrite the result in terms of x. You may also need to use some trigonometric identities and the relationship $\theta = \tan^{-1}\left(\dfrac{x}{a}\right)$. (Note: The reference triangle is based on the assumption that $x > 0$; however, the trigonometric ratios produced from the reference triangle are the same as the ratios for which $x \le 0$.)

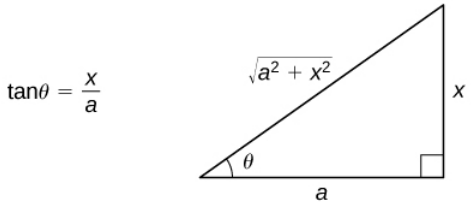


Figure 3.3.4: A reference triangle can be constructed to express the trigonometric functions evaluated at θ in terms of x.

Example 3.3.4: Integrating an Expression Involving $\sqrt{a^2 + x^2}$

Evaluate $\displaystyle\int \frac{dx}{\sqrt{1+x^2}}$ and check the solution by differentiating.

Solution

Begin with the substitution $x = \tan\theta$ and $dx = sec^2\theta\, d\theta$. Since $\tan\theta = x$, draw the reference triangle in Figure 3.3.5.

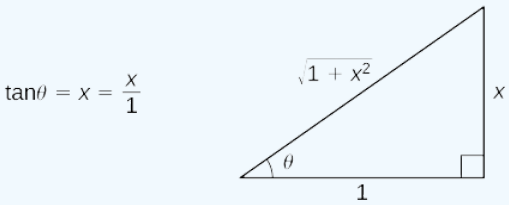


Figure 3.3.5: The reference triangle for Example 3.3.4.

Thus,

$$\int \frac{dx}{\sqrt{1+x^2}} = \int \frac{\sec^2\theta}{\sec\theta}\,d\theta \qquad \text{Substitute } x=\tan\theta \text{ and } dx=\sec^2\theta\,d\theta.$$

$$\text{This substitution makes } \sqrt{1+x^2}=\sec\theta. \text{ Simplify.}$$

$$= \int \sec\theta\,d\theta \qquad \text{Evaluate the integral.}$$

$$= \ln|\sec\theta+\tan\theta|+C \qquad \text{Use the reference triangle to express the result in terms of } x.$$

$$= \ln|\sqrt{1+x^2}+x|+C$$

To check the solution, differentiate:

$$\frac{d}{dx}\left(\ln|\sqrt{1+x^2}+x|\right) = \frac{1}{\sqrt{1+x^2}+x}\cdot\left(\frac{x}{\sqrt{1+x^2}}+1\right) = \frac{1}{\sqrt{1+x^2}+x}\cdot\frac{x+\sqrt{1+x^2}}{\sqrt{1+x^2}} = \frac{1}{\sqrt{1+x^2}}.$$

Since $\sqrt{1+x^2}+x>0$ for all values of x, we could rewrite $\ln|\sqrt{1+x^2}+x|+C=\ln(\sqrt{1+x^2}+x)+C$, if desired.

Example 3.3.5: Evaluating $\int \frac{dx}{\sqrt{1+x^2}}$ Using a Different Substitution

Use the substitution $x=\sinh\theta$ to evaluate $\int \frac{dx}{\sqrt{1+x^2}}$.

Solution

Because $\sinh\theta$ has a range of all real numbers, and $1+\sinh^2\theta=\cosh^2\theta$, we may also use the substitution $x=\sinh\theta$ to evaluate this integral. In this case, $dx=\cosh\theta\,d\theta$. Consequently,

$$\int \frac{dx}{\sqrt{1+x^2}} = \int \frac{\cosh\theta}{\sqrt{1+\sinh^2\theta}}\,d\theta \qquad \text{Substitute } x=\sinh\theta \text{ and } dx=\cosh\theta\,d\theta.$$

$$\text{Substitute } 1+\sinh^2\theta=\cosh^2\theta.$$

$$= \int \frac{\cosh\theta}{\sqrt{\cosh^2\theta}}\,d\theta \qquad \text{Since } \sqrt{\cosh^2\theta}=|\cosh\theta|$$

$$= \int \frac{\cosh\theta}{|\cosh\theta|}\,d\theta \qquad |\cosh\theta|=\cosh\theta \text{ since } \cosh\theta>0 \text{ for all } \theta.$$

$$= \int \frac{\cosh\theta}{\cosh\theta}\,d\theta \qquad \text{Simplify.}$$

$$= \int 1\,d\theta \qquad \text{Evaluate the integral.}$$

$$= \theta+C \qquad \text{Since } x=\sinh\theta, \text{ we know } \theta=\sinh^{-1}x.$$

$$= \sinh^{-1}x+C.$$

Analysis

This answer looks quite different from the answer obtained using the substitution $x=\tan\theta$. To see that the solutions are the same, set $y=\sinh^{-1}x$. Thus, $\sinh y=x$. From this equation we obtain:

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

After multiplying both sides by $2e^y$ and rewriting, this equation becomes:

$$e^{2y}-2xe^y-1=0.$$

Use the quadratic equation to solve for e^y:

$$e^y=\frac{2x\pm\sqrt{4x^2+4}}{2}.$$

Simplifying, we have:

$$e^y = x \pm \sqrt{x^2+1}.$$

Since $x - \sqrt{x^2+1} < 0$, it must be the case that $e^y = x + \sqrt{x^2+1}$. Thus,

$$y = \ln(x + \sqrt{x^2+1}).$$

Last, we obtain

$$\sinh^{-1} x = \ln(x + \sqrt{x^2+1}).$$

After we make the final observation that, since $x + \sqrt{x^2+1} > 0$,

$$\ln(x + \sqrt{x^2+1}) = \ln | \sqrt{1+x^2} + x |,$$

we see that the two different methods produced equivalent solutions.

Example 3.3.6: Finding an Arc Length

Find the length of the curve $y = x^2$ over the interval $[0, \frac{1}{2}]$.

Solution

Because $\frac{dy}{dx} = 2x$, the arc length is given by

$$\int_0^{1/2} \sqrt{1+(2x)^2}\,dx = \int_0^{1/2} \sqrt{1+4x^2}\,dx.$$

To evaluate this integral, use the substitution $x = \frac{1}{2}\tan\theta$ and $dx = \frac{1}{2}\sec^2\theta\,d\theta$. We also need to change the limits of integration. If $x = 0$, then $\theta = 0$ and if $x = \frac{1}{2}$, then $\theta = \frac{\pi}{4}$. Thus,

$\int_0^{1/2} \sqrt{1+4x^2}\,dx = \int_0^{\pi/4} \sqrt{1+\tan^2\theta} \cdot \frac{1}{2}\sec^2\theta\,d\theta$ After substitution, $\sqrt{1+4x^2} = \sec\theta$. (Substitute $1+\tan^2\theta = \sec^2\theta$ and simplify.)

$= \frac{1}{2}\int_0^{\pi/4} \sec^3\theta\,d\theta$ We derived this integral in the previous section.

$= \frac{1}{2}\left(\frac{1}{2}\sec\theta\tan\theta + \frac{1}{2}\ln|\sec\theta+\tan\theta|\right)\Big|_0^{\pi/4}$ Evaluate and simplify.

$= \frac{1}{4}(\sqrt{2} + \ln(\sqrt{2}+1)).$

Exercise 3.3.2

Rewrite $\int x^3\sqrt{x^2+4}\,dx$ by using a substitution involving $\tan\theta$.

Hint

Use $x = 2\tan\theta$ and $dx = 2\sec^2\theta\,d\theta$.

Answer

$$\int 32\tan^3\theta\sec^3\theta\,d\theta$$

Integrating Expressions Involving $\sqrt{x^2 - a^2}$

The domain of the expression $\sqrt{x^2 - a^2}$ is $(-\infty, -a] \cup [a, +\infty)$. Thus, either $x \leq -a$ or $x \geq a$. Hence, $\frac{x}{a} \leq -1$ or $\frac{x}{a} \geq 1$. Since these intervals correspond to the range of $\sec\theta$ on the set $[0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$, it makes sense to use the substitution $\sec\theta = \frac{x}{a}$ or, equivalently, $x = a\sec\theta$, where $0 \leq \theta < \frac{\pi}{2}$ or $\frac{\pi}{2} < \theta \leq \pi$. The corresponding substitution for dx is $dx = a\sec\theta\tan\theta\, d\theta$. The procedure for using this substitution is outlined in the following problem-solving strategy.

Problem-Solving Strategy: Integrals Involving $\sqrt{x^2 - a^2}$

1. Check to see whether the integral cannot be evaluated using another method. If so, we may wish to consider applying an alternative technique.
2. Substitute $x = a\sec\theta$ and $dx = a\sec\theta\tan\theta\, d\theta$. This substitution yields

$$\sqrt{x^2 - a^2} = \sqrt{(a\sec\theta)^2 - a^2} = \sqrt{a^2(\sec^2\theta - 1)} = \sqrt{a^2\tan^2\theta} = |a\tan\theta|.$$

   For $x \geq a$, $|a\tan\theta| = a\tan\theta$ and for $x \leq -a$, $|a\tan\theta| = -a\tan\theta$.
3. Simplify the expression.
4. Evaluate the integral using techniques from the section on trigonometric integrals.
5. Use the reference triangles from Figure 3.3.6 to rewrite the result in terms of x.
6. You may also need to use some trigonometric identities and the relationship $\theta = \sec^{-1}\left(\frac{x}{a}\right)$. (Note: We need both reference triangles, since the values of some of the trigonometric ratios are different depending on whether $x > a$ or $x < -a$.)

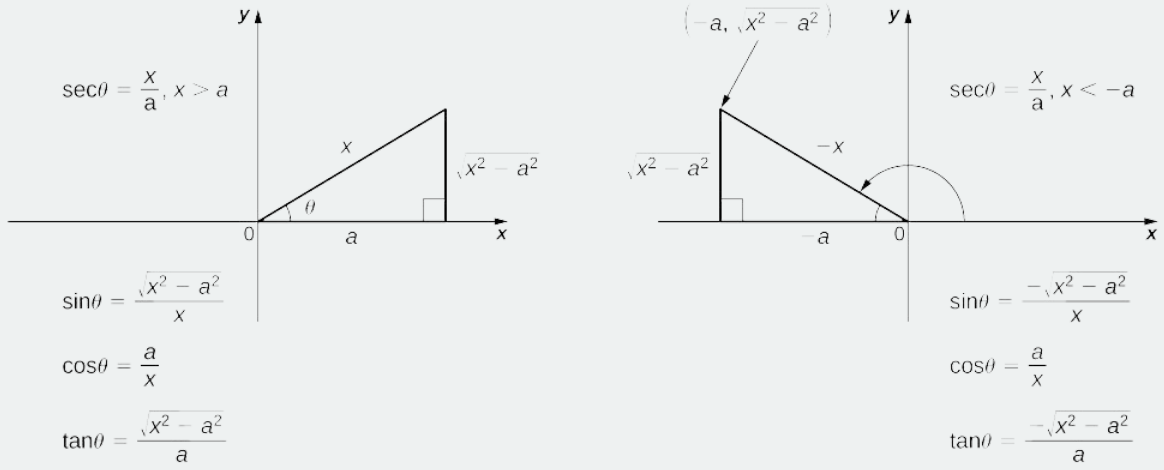


Figure 3.3.6: Use the appropriate reference triangle to express the trigonometric functions evaluated at θ in terms of x.

Example 3.3.7: Finding the Area of a Region

Find the area of the region between the graph of $f(x) = \sqrt{x^2 - 9}$ and the x-axis over the interval $[3, 5]$.

Solution

First, sketch a rough graph of the region described in the problem, as shown in the following figure.

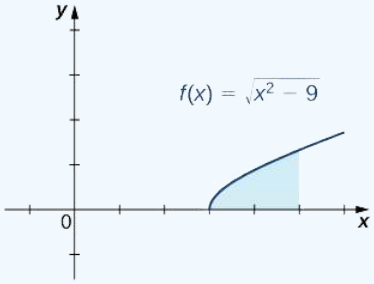


Figure 3.3.7: Calculating the area of the shaded region requires evaluating an integral with a trigonometric substitution.

We can see that the area is $A = \int_3^5 \sqrt{x^2-9}\,dx$. To evaluate this definite integral, substitute $x = 3\sec\theta$ and $dx = 3\sec\theta\tan\theta\,d\theta$. We must also change the limits of integration. If $x = 3$, then $3 = 3\sec\theta$ and hence $\theta = 0$. If $x = 5$, then $\theta = \sec^{-1}(\frac{5}{3})$. After making these substitutions and simplifying, we have

$$\begin{aligned}
\textbf{Area} &= \int_3^5 \sqrt{x^2-9}\,dx \\
&= \int_0^{\sec^{-1}(5/3)} 9\tan^2\theta\sec\theta\,d\theta \quad \text{Use } \tan^2\theta = \sec^2\theta - 1. \\
&= \int_0^{\sec^{-1}(5/3)} 9(\sec^2\theta - 1)\sec\theta\,d\theta \quad \text{Expand.} \\
&= \int_0^{\sec^{-1}(5/3)} 9(\sec^3\theta - \sec\theta)\,d\theta \quad \text{Evaluate the integral.} \\
&= (\frac{9}{2}\ln|\sec\theta+\tan\theta| + \frac{9}{2}\sec\theta\tan\theta) - 9\ln|\sec\theta+\tan\theta|\Big|_0^{\sec^{-1}(5/3)} \quad \text{Simplify.} \\
&= \frac{9}{2}\sec\theta\tan\theta - \frac{9}{2}\ln|\sec\theta+\tan\theta|\Big|_0^{\sec^{-1}(5/3)} \quad \text{Evaluate. Use } \sec(\sec^{-1}\frac{5}{3}) = \frac{5}{3} \text{ and } \tan(\sec^{-1}\frac{5}{3}) = \frac{4}{3}. \\
&= \frac{9}{2}\cdot\frac{5}{3}\cdot\frac{4}{3} - \frac{9}{2}\ln|\frac{5}{3}+\frac{4}{3}| - (\frac{9}{2}\cdot 1\cdot 0 - \frac{9}{2}\ln|1+0|) \\
&= 10 - \frac{9}{2}\ln 3
\end{aligned}$$

Exercise 3.3.3

Evaluate

$$\int \frac{dx}{\sqrt{x^2-4}}.$$

Assume that $x > 2$.

Hint

Substitute $x = 2\sec\theta$ and $dx = 2\sec\theta\tan\theta\,d\theta$.

Answer

$$\ln\left|\frac{x}{2} + \frac{\sqrt{x^2-4}}{2}\right| + C$$

Key Concepts

- For integrals involving $\sqrt{a^2-x^2}$, use the substitution $x = a\sin\theta$ and $dx = a\cos\theta\,d\theta$.
- For integrals involving $\sqrt{a^2+x^2}$, use the substitution $x = a\tan\theta$ and $dx = a\sec^2\theta\,d\theta$.
- For integrals involving $\sqrt{x^2-a^2}$, substitute $x = a\sec\theta$ and $dx = a\sec\theta\tan\theta\,d\theta$.

Glossary

trigonometric substitution

an integration technique that converts an algebraic integral containing expressions of the form $\sqrt{a^2 - x^2}$, $\sqrt{a^2 + x^2}$, or $\sqrt{x^2 - a^2}$ into a trigonometric integral

---

This page titled 3.3: Trigonometric Substitution is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.3: Trigonometric Substitution** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.4: Partial Fractions

Learning Objectives

- Integrate a rational function using the method of partial fractions.
- Recognize simple linear factors in a rational function.
- Recognize repeated linear factors in a rational function.
- Recognize quadratic factors in a rational function.

We have seen some techniques that allow us to integrate specific rational functions. For example, we know that

$$\int \frac{du}{u} = \ln|u| + C$$

and

$$\int \frac{du}{u^2 + a^2} = \frac{1}{a}\tan^{-1}\left(\frac{u}{a}\right) + C.$$

However, we do not yet have a technique that allows us to tackle arbitrary quotients of this type. Thus, it is not immediately obvious how to go about evaluating

$$\int \frac{3x}{x^2 - x - 2}\, dx.$$

However, we know from material previously developed that

$$\int \left(\frac{1}{x+1} + \frac{2}{x-2}\right) dx = \ln|x+1| + 2\ln|x-2| + C.$$

In fact, by getting a common denominator, we see that

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

Consequently,

$$\int \frac{3x}{x^2 - x - 2}\, dx = \int \left(\frac{1}{x+1} + \frac{2}{x-2}\right) dx.$$

In this section, we examine the method of **partial fraction decomposition**, which allows us to decompose **rational functions** into sums of simpler, more easily integrated rational functions. Using this method, we can rewrite an expression such as:

$$\frac{3x}{x^2 - x - 2}$$

as an expression such as

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

The key to the method of partial fraction decomposition is being able to anticipate the form that the decomposition of a rational function will take. As we shall see, this form is both predictable and highly dependent on the factorization of the denominator of the rational function. It is also extremely important to keep in mind that partial fraction decomposition can be applied to a rational function $\frac{P(x)}{Q(x)}$ only if $deg(P(x)) < deg(Q(x))$. In the case when $deg(P(x)) \geq deg(Q(x))$, we must first perform long division to rewrite the quotient $\frac{P(x)}{Q(x)}$ in the form $A(x) + \frac{R(x)}{Q(x)}$, where $deg(R(x)) < deg(Q(x))$. We then do a partial fraction

decomposition on $\frac{R(x)}{Q(x)}$. The following example, although not requiring partial fraction decomposition, illustrates our approach to integrals of rational functions of the form $\int \frac{P(x)}{Q(x)}\,dx$, where $deg(P(x)) \geq deg(Q(x))$.

Example 3.4.1: Integrating $\int \frac{P(x)}{Q(x)}\,dx$, where $deg(P(x)) \geq deg(Q(x))$

Evaluate

$$\int \frac{x^2+3x+5}{x+1}\,dx.$$

Solution

Since $deg(x^2+3x+5) \geq deg(x+1)$, we perform long division to obtain

$$\frac{x^2+3x+5}{x+1} = x+2+\frac{3}{x+1}.$$

Thus,

$$\int \frac{x^2+3x+5}{x+1}\,dx = \int \left(x+2+\frac{3}{x+1}\right)\,dx = \frac{1}{2}x^2+2x+3\ln|x+1|+C.$$

Visit this website for a review of long division of polynomials.

Exercise 3.4.1

Evaluate

$$\int \frac{x-3}{x+2}\,dx.$$

Hint

Use long division to obtain $\frac{x-3}{x+2} = 1-\frac{5}{x+2}$.

Answer

$$x-5\ln|x+2|+C$$

To integrate $\int \frac{P(x)}{Q(x)}\,dx$, where $deg(P(x)) < deg(Q(x))$, we must begin by factoring $Q(x)$.

Nonrepeated Linear Factors

If $Q(x)$ can be factored as $(a_1x+b_1)(a_2x+b_2)\ldots(a_nx+b_n)$, where each linear factor is distinct, then it is possible to find constants $A_1, A_2, \ldots A_n$ satisfying

$$\frac{P(x)}{Q(x)} = \frac{A_1}{a_1x+b_1}+\frac{A_2}{a_2x+b_2}+\cdots+\frac{A_n}{a_nx+b_n}. \tag{3.4.1}$$

The proof that such constants exist is beyond the scope of this course.

In this next example, we see how to use partial fractions to integrate a rational function of this type.

Example 3.4.2: Partial Fractions with Nonrepeated Linear Factors

Evaluate $\int \frac{3x+2}{x^3-x^2-2x}\,dx$.

Solution

Since $deg(3x+2) < deg(x^3-x^2-2x)$, we begin by factoring the denominator of $\frac{3x+2}{x^3-x^2-2x}$. We can see that $x^3-x^2-2x = x(x-2)(x+1)$. Thus, there are constants A, B, and C satisfying Equation 3.4.1 such that

$$\frac{3x+2}{x(x-2)(x+1)} = \frac{A}{x} + \frac{B}{x-2} + \frac{C}{x+1}.$$

We must now find these constants. To do so, we begin by getting a common denominator on the right. Thus,

$$\frac{3x+2}{x(x-2)(x+1)} = \frac{A(x-2)(x+1)+Bx(x+1)+Cx(x-2)}{x(x-2)(x+1)}.$$

Now, we set the numerators equal to each other, obtaining

$$3x+2 = A(x-2)(x+1)+Bx(x+1)+Cx(x-2). \tag{3.4.2}$$

There are two different strategies for finding the coefficients A, B, and C. We refer to these as the *method of equating coefficients* and the *method of strategic substitution*.

Strategy one: Method of Equating Coefficients

Rewrite Equation 3.4.2 in the form

$$3x+2 = (A+B+C)x^2+(-A+B-2C)x+(-2A).$$

Equating coefficients produces the system of equations

$$\begin{aligned} A+B+C &= 0 \\ -A+B-2C &= 3 \\ -2A &= 2. \end{aligned}$$

To solve this system, we first observe that $-2A = 2 \Rightarrow A = -1$. Substituting this value into the first two equations gives us the system

$$\begin{aligned} B+C &= 1 \\ B-2C &= 2\,. \end{aligned}$$

Multiplying the second equation by -1 and adding the resulting equation to the first produces

$$-3C = 1,$$

which in turn implies that $C = -\frac{1}{3}$. Substituting this value into the equation $B+C=1$ yields $B = \frac{4}{3}$. Thus, solving these equations yields $A = -1, B = \frac{4}{3}$, and $C = -\frac{1}{3}$.

It is important to note that the system produced by this method is consistent if and only if we have set up the decomposition correctly. If the system is inconsistent, there is an error in our decomposition.

Strategy two: Method of Strategic Substitution

The method of strategic substitution is based on the assumption that we have set up the decomposition correctly. If the decomposition is set up correctly, then there must be values of A, B, and C that satisfy Equation 3.4.2 for all values of x. That is, this equation must be true for any value of x we care to substitute into it. Therefore, by choosing values of x carefully and substituting them into the equation, we may find A, B, and C easily. For example, if we substitute $x = 0$, the equation reduces to $2 = A(-2)(1)$. Solving for A yields $A = -1$. Next, by substituting $x = 2$, the equation reduces to $8 = B(2)(3)$, or

equivalently $B=4/3$. Last, we substitute $x=-1$ into the equation and obtain $-1=C(-1)(-3)$. Solving, we have $C=-\frac{1}{3}$.

It is important to keep in mind that if we attempt to use this method with a decomposition that has not been set up correctly, we are still able to find values for the constants, but these constants are meaningless. If we do opt to use the method of strategic substitution, then it is a good idea to check the result by recombining the terms algebraically.

Now that we have the values of A, B, and C, we rewrite the original integral:

$$\int \frac{3x+2}{x^3-x^2-2x}\,dx=\int\left(-\frac{1}{x}+\frac{4}{3}\cdot\frac{1}{x-2}-\frac{1}{3}\cdot\frac{1}{x+1}\right)\,dx.$$

Evaluating the integral gives us

$$\int \frac{3x+2}{x^3-x^2-2x}\,dx=-\ln|x|+\frac{4}{3}\ln|x-2|-\frac{1}{3}\ln|x+1|+C.$$

In the next example, we integrate a rational function in which the degree of the numerator is not less than the degree of the denominator.

Example 3.4.3: Dividing before Applying Partial Fractions

Evaluate $\int \frac{x^2+3x+1}{x^2-4}\,dx$.

Solution

Since $deg(x^2+3x+1)\geq deg(x^2-4)$, we must perform long division of polynomials. This results in

$$\frac{x^2+3x+1}{x^2-4}=1+\frac{3x+5}{x^2-4}$$

Next, we perform partial fraction decomposition on $\frac{3x+5}{x^2-4}=\frac{3x+5}{(x+2)(x-2)}$. We have

$$\frac{3x+5}{(x-2)(x+2)}=\frac{A}{x-2}+\frac{B}{x+2}.$$

Thus,

$$3x+5=A(x+2)+B(x-2).$$

Solving for A and B using either method, we obtain $A=11/4$ and $B=1/4$.

Rewriting the original integral, we have

$$\int \frac{x^2+3x+1}{x^2-4}\,dx=\int\left(1+\frac{11}{4}\cdot\frac{1}{x-2}+\frac{1}{4}\cdot\frac{1}{x+2}\right)\,dx.$$

Evaluating the integral produces

$$\int \frac{x^2+3x+1}{x^2-4}\,dx=x+\frac{11}{4}\ln|x-2|+\frac{1}{4}\ln|x+2|+C.$$

As we see in the next example, it may be possible to apply the technique of partial fraction decomposition to a nonrational function. The trick is to convert the nonrational function to a rational function through a substitution.

Example 3.4.4: Applying Partial Fractions after a Substitution

Evaluate $\int \frac{\cos x}{\sin^2 x-\sin x}\,dx$.

Solution

Let's begin by letting $u = \sin x$. Consequently, $du = \cos x\, dx$. After making these substitutions, we have

$$\int \frac{\cos x}{\sin^2 x - \sin x}\, dx = \int \frac{du}{u^2 - u} = \int \frac{du}{u(u-1)}.$$

Applying partial fraction decomposition to $\frac{1}{u(u-1)}$ gives $\frac{1}{u(u-1)} = -\frac{1}{u} + \frac{1}{u-1}$.

Thus,

$$\int \frac{\cos x}{\sin^2 x - \sin x}\, dx = -\ln|u| + \ln|u-1| + C = -\ln|\sin x| + \ln|\sin x - 1| + C.$$

Exercise 3.4.2

Evaluate $\int \frac{x+1}{(x+3)(x-2)}\, dx$.

Hint

$$\frac{x+1}{(x+3)(x-2)} = \frac{A}{x+3} + \frac{B}{x-2}$$

Answer

$$\frac{2}{5}\ln|x+3| + \frac{3}{5}\ln|x-2| + C$$

Repeated Linear Factors

For some applications, we need to integrate rational expressions that have denominators with repeated linear factors—that is, rational functions with at least one factor of the form $(ax+b)^n$, where n is a positive integer greater than or equal to 2. If the denominator contains the repeated linear factor $(ax+b)^n$, then the decomposition must contain

$$\frac{A_1}{ax+b} + \frac{A_2}{(ax+b)^2} + \cdots + \frac{A_n}{(ax+b)^n}. \tag{3.4.3}$$

As we see in our next example, the basic technique used for solving for the coefficients is the same, but it requires more algebra to determine the numerators of the partial fractions.

Example 3.4.5: Partial Fractions with Repeated Linear Factors

Evaluate $\int \frac{x-2}{(2x-1)^2(x-1)}\, dx$.

Solution

We have $deg(x-2) < deg((2x-1)^2(x-1))$, so we can proceed with the decomposition. Since $(2x-1)^2$ is a repeated linear factor, include

$$\frac{A}{2x-1} + \frac{B}{(2x-1)^2}$$

in the decomposition in Equation 3.4.3. Thus,

$$\frac{x-2}{(2x-1)^2(x-1)} = \frac{A}{2x-1} + \frac{B}{(2x-1)^2} + \frac{C}{x-1}.$$

After getting a common denominator and equating the numerators, we have

$$x - 2 = A(2x-1)(x-1) + B(x-1) + C(2x-1)^2. \tag{3.4.4}$$

We then use the method of equating coefficients to find the values of A, B, and C.

$$x-2=(2A+4C)x^2+(-3A+B-4C)x+(A-B+C).$$

Equating coefficients yields $2A+4C=0$, $-3A+B-4C=1$, and $A-B+C=-2$. Solving this system yields $A=2, B=3,$ and $C=-1$.

Alternatively, we can use the method of strategic substitution. In this case, substituting $x=1$ and $x=1/2$ into Equation 3.4.4 easily produces the values $B=3$ and $C=-1$. At this point, it may seem that we have run out of good choices for x, however, since we already have values for B and C, we can substitute in these values and choose any value for x not previously used. The value $x=0$ is a good option. In this case, we obtain the equation $-2=A(-1)(-1)+3(-1)+(-1)(-1)^2$ or, equivalently, $A=2$.

Now that we have the values for A, B, and C, we rewrite the original integral and evaluate it:

$$\begin{aligned}\int \frac{x-2}{(2x-1)^2(x-1)}\,dx &= \int\left(\frac{2}{2x-1}+\frac{3}{(2x-1)^2}-\frac{1}{x-1}\right)dx\\ &= \ln|2x-1|-\frac{3}{2(2x-1)}-\ln|x-1|+C.\end{aligned}$$

? Exercise 3.4.3

Set up the partial fraction decomposition for

$$\int \frac{x+2}{(x+3)^3(x-4)^2}\,dx.$$

(Do not solve for the coefficients or complete the integration.)

Hint

Use the problem-solving method of Example 3.4.5for guidance.

Answer

$$\frac{x+2}{(x+3)^3(x-4)^2}=\frac{A}{x+3}+\frac{B}{(x+3)^2}+\frac{C}{(x+3)^3}+\frac{D}{(x-4)}+\frac{E}{(x-4)^2}$$

The General Method

Now that we are beginning to get the idea of how the technique of partial fraction decomposition works, let's outline the basic method in the following problem-solving strategy.

Problem-Solving Strategy: Partial Fraction Decomposition

To decompose the rational function $P(x)/Q(x)$, use the following steps:

1. Make sure that $deg(P(x))<deg(Q(x))$. If not, perform long division of polynomials.
2. Factor $Q(x)$ into the product of linear and irreducible quadratic factors. An irreducible quadratic is a quadratic that has no real zeros.
3. Assuming that $deg(P(x))<deg(Q(x)$, the factors of $Q(x)$ determine the form of the decomposition of $P(x)/Q(x)$.
   a. If $Q(x)$ can be factored as $(a_1x+b_1)(a_2x+b_2)\ldots(a_nx+b_n)$, where each linear factor is distinct, then it is possible to find constants $A_1, A_2, \ldots A_n$ satisfying

$$\frac{P(x)}{Q(x)}=\frac{A_1}{a_1x+b_1}+\frac{A_2}{a_2x+b_2}+\cdots+\frac{A_n}{a_nx+b_n}.$$

   b. If $Q(x)$ contains the repeated linear factor $(ax+b)^n$, then the decomposition must contain

$$\frac{A_1}{ax+b}+\frac{A_2}{(ax+b)^2}+\cdots+\frac{A_n}{(ax+b)^n}.$$

c. For each irreducible quadratic factor ax^2+bx+c that $Q(x)$ contains, the decomposition must include

$$\frac{Ax+B}{ax^2+bx+c}.$$

d. For each repeated irreducible quadratic factor $(ax^2+bx+c)^n$, the decomposition must include

$$\frac{A_1x+B_1}{ax^2+bx+c}+\frac{A_2x+B_2}{(ax^2+bx+c)^2}+\cdots+\frac{A_nx+B_n}{(ax^2+bx+c)^n}.$$

e. After the appropriate decomposition is determined, solve for the constants.

f. Last, rewrite the integral in its decomposed form and evaluate it using previously developed techniques or integration formulas.

Simple Quadratic Factors

Now let's look at integrating a rational expression in which the denominator contains an irreducible quadratic factor. Recall that the quadratic ax^2+bx+c is irreducible if $ax^2+bx+c=0$ has no real zeros—that is, if $b^2-4ac<0$.

Example 3.4.6: Rational Expressions with an Irreducible Quadratic Factor

Evaluate

$$\int\frac{2x-3}{x^3+x}\,dx.$$

Solution

Since $deg(2x-3)<deg(x^3+x)$, factor the denominator and proceed with partial fraction decomposition. Since $x^3+x=x(x^2+1)$ contains the irreducible quadratic factor x^2+1, include $\frac{Ax+B}{x^2+1}$ as part of the decomposition, along with $\frac{C}{x}$ for the linear term x. Thus, the decomposition has the form

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

After getting a common denominator and equating the numerators, we obtain the equation

$$2x-3=(Ax+B)x+C(x^2+1).$$

Solving for A, B, and C, we get $A=3$, $B=2$, and $C=-3$.

Thus,

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

Substituting back into the integral, we obtain

$$\begin{aligned}\int\frac{2x-3}{x^3+x}\,dx &= \int\left(\frac{3x+2}{x^2+1}-\frac{3}{x}\right)dx \\ &= 3\int\frac{x}{x^2+1}\,dx+2\int\frac{1}{x^2+1}\,dx-3\int\frac{1}{x}\,dx && \text{Split up the integral} \\ &= \frac{3}{2}\ln|x^2+1|+2\tan^{-1}x-3\ln|x|+C. && \text{Evaluate each integral}\end{aligned}$$

Note: We may rewrite $\ln|x^2+1|=\ln(x^2+1)$, if we wish to do so, since $x^2+1>0$.

Example 3.4.7: Partial Fractions with an Irreducible Quadratic Factor

Evaluate $\int \frac{dx}{x^3-8}$.

Solution: We can start by factoring $x^3-8=(x-2)(x^2+2x+4)$. We see that the quadratic factor x^2+2x+4 is irreducible since $2^2-4(1)(4)=-12<0$. Using the decomposition described in the problem-solving strategy, we get

$$\frac{1}{(x-2)(x^2+2x+4)}=\frac{A}{x-2}+\frac{Bx+C}{x^2+2x+4}.$$

After obtaining a common denominator and equating the numerators, this becomes

$$1=A(x^2+2x+4)+(Bx+C)(x-2).$$

Applying either method, we get $A=\frac{1}{12}, B=-\frac{1}{12},$ and $C=-\frac{1}{3}$.

Rewriting $\int \frac{dx}{x^3-8}$, we have

$$\int \frac{dx}{x^3-8}=\frac{1}{12}\int \frac{1}{x-2}\,dx-\frac{1}{12}\int \frac{x+4}{x^2+2x+4}\,dx.$$

We can see that

$$\int \frac{1}{x-2}\,dx=\ln|x-2|+C,$$

but

$$\int \frac{x+4}{x^2+2x+4}\,dx$$

requires a bit more effort. Let's begin by **completing the square** on x^2+2x+4 to obtain

$$x^2+2x+4=(x+1)^2+3.$$

By letting $u=x+1$ and consequently $du=dx$, we see that

$$\begin{aligned}
\int \frac{x+4}{x^2+2x+4}\,dx &= \int \frac{x+4}{(x+1)^2+3}\,dx && \text{Complete the square on the denominator}\\
&= \int \frac{u+3}{u^2+3}\,du && \text{Substitute } u=x+1,\ x=u-1, \text{ and } du=dx\\
&= \int \frac{u}{u^2+3}du+\int \frac{3}{u^2+3}du && \text{Split the numerator apart}\\
&= \frac{1}{2}\ln|u^2+3|+\frac{3}{\sqrt{3}}\tan^{-1}\frac{u}{\sqrt{3}}+C && \text{Evaluate each integral}\\
&= \frac{1}{2}\ln|x^2+2x+4|+\sqrt{3}\tan^{-1}\left(\frac{x+1}{\sqrt{3}}\right)+C && \text{Rewrite in terms of } x \text{ and simplify}
\end{aligned}$$

Substituting back into the original integral and simplifying gives

$$\int \frac{dx}{x^3-8}=\frac{1}{12}\ln|x-2|-\frac{1}{24}\ln|x^2+2x+4|-\frac{\sqrt{3}}{12}\tan^{-1}\left(\frac{x+1}{\sqrt{3}}\right)+C.$$

Here again, we can drop the absolute value if we wish to do so, since $x^2+2x+4>0$ for all x.

Example 3.4.8: Finding a Volume

Find the volume of the solid of revolution obtained by revolving the region enclosed by the graph of $f(x)=\dfrac{x^2}{(x^2+1)^2}$ and the x-axis over the interval $[0,1]$ about the y-axis.

Solution

Let's begin by sketching the region to be revolved (see Figure 3.4.1). From the sketch, we see that the shell method is a good choice for solving this problem.

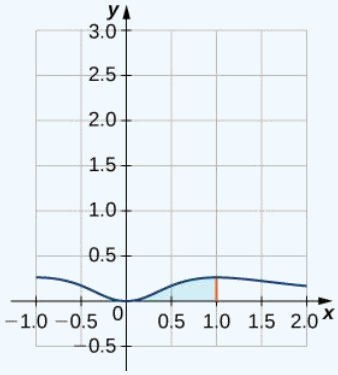


Figure 3.4.1: We can use the shell method to find the volume of revolution obtained by revolving the region shown about the y-axis.

The volume is given by

$$V=2\pi\int_0^1 x\cdot\frac{x^2}{(x^2+1)^2}\,dx=2\pi\int_0^1\frac{x^3}{(x^2+1)^2}\,dx.$$

Since $deg((x^2+1)^2)=4>3=deg(x^3),$ we can proceed with partial fraction decomposition. Note that $(x^2+1)^2$ is a repeated irreducible quadratic. Using the decomposition described in the problem-solving strategy, we get

$$\frac{x^3}{(x^2+1)^2}=\frac{Ax+B}{x^2+1}+\frac{Cx+D}{(x^2+1)^2}.$$

Finding a common denominator and equating the numerators gives

$$x^3=(Ax+B)(x^2+1)+Cx+D.$$

Solving, we obtain $A=1, B=0, C=-1,$ and $D=0$. Substituting back into the integral, we have

$$V=2\pi\int_0^1\frac{x^3}{(x^2+1)^2}\,dx=2\pi\int_0^1\left(\frac{x}{x^2+1}-\frac{x}{(x^2+1)^2}\right)dx=2\pi\left(\frac{1}{2}\ln(x^2+1)+\frac{1}{2}\cdot\frac{1}{x^2+1}\right)\Big|_0^1=$$
$$\pi\left(\ln 2-\tfrac{1}{2}\right).$$

Exercise 3.4.4

Set up the partial fraction decomposition for

$$\int\frac{x^2+3x+1}{(x+2)(x-3)^2(x^2+4)^2}\,dx.$$

Hint

Use the problem-solving strategy.

Answer

$$\frac{x^2+3x+1}{(x+2)(x-3)^2(x^2+4)^2}=\frac{A}{x+2}+\frac{B}{x-3}+\frac{C}{(x-3)^2}+\frac{Dx+E}{x^2+4}+\frac{Fx+G}{(x^2+4)^2}$$

Key Concepts

- Partial fraction decomposition is a technique used to break down a rational function into a sum of simple rational functions that can be integrated using previously learned techniques.
- When applying partial fraction decomposition, we must make sure that the degree of the numerator is less than the degree of the denominator. If not, we need to perform long division before attempting partial fraction decomposition.
- The form the decomposition takes depends on the type of factors in the denominator. The types of factors include nonrepeated linear factors, repeated linear factors, nonrepeated irreducible quadratic factors, and repeated irreducible quadratic factors.

Glossary

partial fraction decomposition

a technique used to break down a rational function into the sum of simple rational functions

---

This page titled 3.4: Partial Fractions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.4: Partial Fractions** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.5: Other Strategies for Integration

Learning Objectives

- Use a table of integrals to solve integration problems.
- Use a computer algebra system (CAS) to solve integration problems.

In addition to the techniques of integration we have already seen, several other tools are widely available to assist with the process of integration. Among these tools are **integration tables**, which are readily available in many books, including the appendices to this one. Also widely available are **computer algebra systems (CAS)**, which are found on calculators and in many campus computer labs, and are free online.

Tables of Integrals

Integration tables, if used in the right manner, can be a handy way either to evaluate or check an integral quickly. Keep in mind that when using a table to check an answer, it is possible for two completely correct solutions to look very different. For example, in Trigonometric Substitution, we found that, by using the substitution $x = \tan\theta$, we can arrive at

$$\int \frac{dx}{\sqrt{1+x^2}} = \ln\left|x+\sqrt{x^2+1}\right| + C.$$

However, using $x = \sinh\theta$, we obtained a different solution—namely,

$$\int \frac{dx}{\sqrt{1+x^2}} = \sinh^{-1} x + C.$$

We later showed algebraically that the two solutions are equivalent. That is, we showed that $\sinh^{-1} x = \ln\left|x+\sqrt{x^2+1}\right|$. In this case, the two antiderivatives that we found were actually equal. This need not be the case. However, as long as the difference in the two antiderivatives is a constant, they are equivalent.

Example 3.5.1: Using a Formula from a Table to Evaluate an Integral

Use the table formula

$$\int \frac{\sqrt{a^2-u^2}}{u^2}du = -\frac{\sqrt{a^2-u^2}}{u} - \sin^{-1}\frac{u}{a} + C$$

to evaluate $\displaystyle\int \frac{\sqrt{16-e^{2x}}}{e^x}dx$.

Solution

If we look at integration tables, we see that several formulas contain expressions of the form $\sqrt{a^2-u^2}$. This expression is actually similar to $\sqrt{16-e^{2x}}$, where $a=4$ and $u=e^x$. Keep in mind that we must also have $du=e^x$. Multiplying the numerator and the denominator of the given integral by e^x should help to put this integral in a useful form. Thus, we now have

$$\int \frac{\sqrt{16-e^{2x}}}{e^x}dx = \int \frac{\sqrt{16-e^{2x}}}{e^{2x}}e^x dx.$$

Substituting $u=e^x$ and $du=e^x\,dx$ produces $\displaystyle\int \frac{\sqrt{a^2-u^2}}{u^2}du$. From the integration table (#88 in Appendix A),

$$\int \frac{\sqrt{a^2-u^2}}{u^2}du = -\frac{\sqrt{a^2-u^2}}{u} - \sin^{-1}\frac{u}{a} + C.$$

Thus,

$$\begin{aligned}\int \frac{\sqrt{16-e^{2x}}}{e^x}dx &= \int \frac{\sqrt{16-e^{2x}}}{e^{2x}}e^x dx \text{ Substitute } u=e^x \text{ and } du=e^x dx.\\ &= \int \frac{\sqrt{4^2-u^2}}{u^2}du \text{ Apply the formula using } a=4.\end{aligned}$$

$$= -\frac{\sqrt{4^2-u^2}}{u} - \sin^{-1}\frac{u}{4} + C \quad \text{Substitute } u = e^x.$$
$$= -\frac{\sqrt{16-e^{2x}}}{e^x} - \sin^{-1}\left(\frac{e^x}{4}\right) + C$$

Computer Algebra Systems

If available, a CAS is a faster alternative to a table for solving an integration problem. Many such systems are widely available and are, in general, quite easy to use.

Example 3.5.2: Using a Computer Algebra System to Evaluate an Integral

Use a computer algebra system to evaluate $\int \frac{dx}{\sqrt{x^2-4}}$. Compare this result with $\ln\left|\frac{\sqrt{x^2-4}}{2} + \frac{x}{2}\right| + C$, a result we might have obtained if we had used trigonometric substitution.

Solution

Using Wolfram Alpha, we obtain

$$\int \frac{dx}{\sqrt{x^2-4}} = \ln\left|\sqrt{x^2-4} + x\right| + C.$$

Notice that

$$\ln\left|\frac{\sqrt{x^2-4}}{2} + \frac{x}{2}\right| + C = \ln\left|\frac{\sqrt{x^2-4}+x}{2}\right| + C = \ln\left|\sqrt{x^2-4}+x\right| - \ln 2 + C.$$

Since these two antiderivatives differ by only a constant, the solutions are equivalent. We could have also demonstrated that each of these antiderivatives is correct by differentiating them.

You can access an integral calculator for more examples.

Example 3.5.3: Using a CAS to Evaluate an Integral

Evaluate $\int \sin^3 x\, dx$ using a CAS. Compare the result to $\frac{1}{3}\cos^3 x - \cos x + C$, the result we might have obtained using the technique for integrating odd powers of $\sin x$ discussed earlier in this chapter.

Solution

Using Wolfram Alpha, we obtain

$$\int \sin^3 x\, dx = \frac{1}{12}(\cos(3x) - 9\cos x) + C.$$

This looks quite different from $\frac{1}{3}\cos^3 x - \cos x + C$. To see that these antiderivatives are equivalent, we can make use of a few trigonometric identities:

$$\begin{aligned}
\frac{1}{12}(\cos(3x) - 9\cos x) &= \frac{1}{12}(\cos(x+2x) - 9\cos x) \\
&= \frac{1}{12}(\cos(x)\cos(2x) - \sin(x)\sin(2x) - 9\cos x) \\
&= \frac{1}{12}(\cos x(2\cos^2 x - 1) - \sin x(2\sin x\cos x) - 9\cos x) \\
&= \frac{1}{12}(2\cos^3 x - \cos x - 2\cos x(1-\cos^2 x) - 9\cos x) \\
&= \frac{1}{12}(4\cos^3 x - 12\cos x)
\end{aligned}$$

$$= \frac{1}{3}\cos^3 x - \cos x.$$

Thus, the two antiderivatives are identical.

We may also use a CAS to compare the graphs of the two functions, as shown in the following figure.

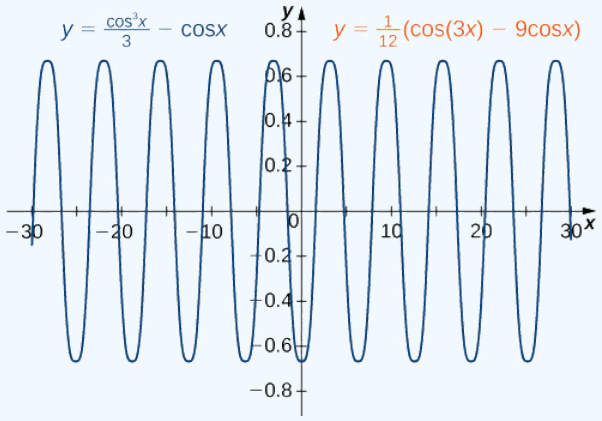


Figure 3.5.1: The graphs of $y = \frac{1}{3}\cos^3 x - \cos x$ and $y = \frac{1}{12}(\cos(3x) - 9\cos x)$ are identical.

Exercise 3.5.1

Use a CAS to evaluate $\int \frac{dx}{\sqrt{x^2+4}}$.

Hint

Answers may vary.

Answer

Possible solutions include $\sinh^{-1}\left(\frac{x}{2}\right) + C$ and $\ln\left|\sqrt{x^2+4} + x\right| + C$.

Key Concepts

- An integration table may be used to evaluate indefinite integrals.
- A CAS (or computer algebra system) may be used to evaluate indefinite integrals.
- It may require some effort to reconcile equivalent solutions obtained using different methods.

Glossary

computer algebra system (CAS)

technology used to perform many mathematical tasks, including integration

integration table

a table that lists integration formulas

This page titled 3.5: Other Strategies for Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.5: Other Strategies for Integration** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.6: Numerical Integration

Learning Objectives

- Approximate the value of a definite integral by using the midpoint and trapezoidal rules.
- Determine the absolute and relative error in using a numerical integration technique.
- Estimate the absolute and relative error using an error-bound formula.
- Recognize when the midpoint and trapezoidal rules over- or underestimate the true value of an integral.
- Use Simpson's rule to approximate the value of a definite integral to a given accuracy.

The antiderivatives of many functions either cannot be expressed or cannot be expressed easily in closed form (that is, in terms of known functions). Consequently, rather than evaluate definite integrals of these functions directly, we resort to various techniques of **numerical integration** to approximate their values. In this section, we explore several of these techniques. In addition, we examine the process of estimating the error in using these techniques.

The Midpoint Rule

Earlier in this text we defined the definite integral of a function over an interval as the limit of **Riemann sums**. In general, any Riemann sum of a function $f(x)$ over an interval $[a, b]$ may be viewed as an estimate of $\int_a^b f(x)\,dx$. Recall that a Riemann sum of a function $f(x)$ over an interval $[a, b]$ is obtained by selecting a partition

$$P = \{x_0, x_1, x_2, \ldots, x_n\}$$

where $a = x_0 < x_1 < x_2 < \cdots < x_n = b$

and a set

$$S = \{x_1^*, x_2^*, \ldots, x_n^*\}$$

where $x_{i-1} \le x_i^* \le x_i \quad$ for all i.

The Riemann sum corresponding to the partition P and the set S is given by $\sum_{i=1}^{n} f(x_i^*)\Delta x_i$, where $\Delta x_i = x_i - x_{i-1}$, the length of the ith subinterval.

The midpoint rule for estimating a definite integral uses a Riemann sum with subintervals of equal width and the midpoints, m_i, of each subinterval in place of x_i^*. Formally, we state a theorem regarding the convergence of the midpoint rule as follows.

The Midpoint Rule

Assume that $f(x)$ is continuous on $[a, b]$. Let n be a positive integer and $\Delta x = \dfrac{b-a}{n}$. If $[a, b]$ is divided into n subintervals, each of length Δx, and m_i is the midpoint of the ith subinterval, set

$$M_n = \sum_{i=1}^{n} f(m_i)\Delta x.$$

Then $\lim_{n\to\infty} M_n = \int_a^b f(x)\,dx$.

As we can see in Figure 3.6.1, if $f(x) \ge 0$ over $[a, b]$, then $\sum_{i=1}^{n} f(m_i)\Delta x$ corresponds to the sum of the areas of rectangles approximating the area between the graph of $f(x)$ and the x-axis over $[a, b]$. The graph shows the rectangles corresponding to M_4 for a nonnegative function over a closed interval $[a, b]$.

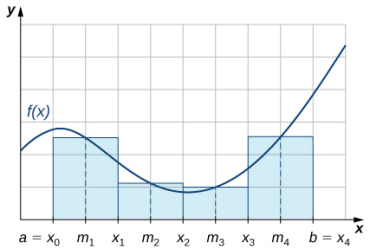


Figure 3.6.1: The midpoint rule approximates the area between the graph of $f(x)$ and the x-axis by summing the areas of rectangles with midpoints that are points on $f(x)$.

Example 3.6.1: Using the Midpoint Rule with M_4

Use the midpoint rule to estimate $\int_0^1 x^2\,dx$ using four subintervals. Compare the result with the actual value of this integral.

Solution: Each subinterval has length $\Delta x = \dfrac{1-0}{4} = \dfrac{1}{4}$. Therefore, the subintervals consist of

$$\left[0, \tfrac{1}{4}\right], \left[\tfrac{1}{4}, \tfrac{1}{2}\right], \left[\tfrac{1}{2}, \tfrac{3}{4}\right], \text{ and } \left[\tfrac{3}{4}, 1\right].$$

The midpoints of these subintervals are $\left\{\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}\right\}$. Thus,

$$\begin{aligned} M_4 &= \frac{1}{4}\cdot f\left(\frac{1}{8}\right)+\frac{1}{4}\cdot f\left(\frac{3}{8}\right)+\frac{1}{4}\cdot f\left(\frac{5}{8}\right)+\frac{1}{4}\cdot f\left(\frac{7}{8}\right) \\ &= \frac{1}{4}\cdot\frac{1}{64}+\frac{1}{4}\cdot\frac{9}{64}+\frac{1}{4}\cdot\frac{25}{64}+\frac{1}{4}\cdot\frac{49}{64} \\ &= \frac{21}{64}=0.328125. \end{aligned}$$

Since

$$\int_0^1 x^2\,dx=\frac{1}{3},$$

the absolute error in this approximation is:

$$\left|\frac{1}{3}-\frac{21}{64}\right|=\frac{1}{192}\approx 0.0052,$$

and we see that the midpoint rule produces an estimate that is somewhat close to the actual value of the definite integral.

Example 3.6.2: Using the Midpoint Rule with M_6

Use M_6 to estimate the length of the curve $y=\frac{1}{2}x^2$ on $[1,4]$.

Solution: The length of $y=\frac{1}{2}x^2$ on $[1,4]$ is

$$s=\int_1^4\sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx.$$

Since $\frac{dy}{dx}=x$, this integral becomes $\int_1^4\sqrt{1+x^2}\,dx$.

If $[1,4]$ is divided into six subintervals, then each subinterval has length $\Delta x=\frac{4-1}{6}=\frac{1}{2}$ and the midpoints of the subintervals are $\left\{\frac{5}{4},\frac{7}{4},\frac{9}{4},\frac{11}{4},\frac{13}{4},\frac{15}{4}\right\}$. If we set $f(x)=\sqrt{1+x^2}$,

$$\begin{aligned} M_6 &= \tfrac{1}{2}\cdot f\left(\frac{5}{4}\right)+\tfrac{1}{2}\cdot f\left(\frac{7}{4}\right)+\frac{1}{2}\cdot f\left(\frac{9}{4}\right)+\frac{1}{2}\cdot f\left(\frac{11}{4}\right)+\frac{1}{2}\cdot f\left(\frac{13}{4}\right)+\frac{1}{2}\cdot f\left(\frac{15}{4}\right) \\ &\approx \frac{1}{2}(1.6008+2.0156+2.4622+2.9262+3.4004+3.8810)=8.1431 \text{ units.} \end{aligned}$$

Exercise 3.6.1

Use the midpoint rule with $n=2$ to estimate $\int_1^2\frac{1}{x}\,dx$.

Hint

$\Delta x=\frac{1}{2}$, $m_1=\frac{5}{4}$, and $m_2=\frac{7}{4}$.

Answer

$\frac{24}{35}\approx 0.685714$

The Trapezoidal Rule

We can also approximate the value of a definite integral by using trapezoids rather than rectangles. In Figure 3.6.2, the area beneath the curve is approximated by trapezoids rather than by rectangles.

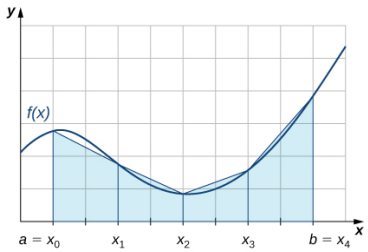


Figure 3.6.2: Trapezoids may be used to approximate the area under a curve, hence approximating the definite integral.

The **trapezoidal rule** for estimating definite integrals uses trapezoids rather than rectangles to approximate the area under a curve. To gain insight into the final form of the rule, consider the trapezoids shown in Figure 3.6.2. We assume that the length of each subinterval is given by Δx. First, recall that the area of a trapezoid with a height of h and bases of length b_1 and b_2 is given by $\text{Area}=\frac{1}{2}h(b_1+b_2)$. We see that the first trapezoid has a height Δx and parallel bases of length $f(x_0)$ and $f(x_1)$. Thus, the area of the first trapezoid in Figure 3.6.2 is

$$\frac{1}{2}\Delta x\Big(f(x_0)+f(x_1)\Big).$$

The areas of the remaining three trapezoids are

$$\frac{1}{2}\Delta x\Big(f(x_1)+f(x_2)\Big),\ \frac{1}{2}\Delta x\Big(f(x_2)+f(x_3)\Big),\text{ and }\frac{1}{2}\Delta x\Big(f(x_3)+f(x_4)\Big).$$

Consequently,

$$\int_a^b f(x)\,dx \approx \frac{1}{2}\Delta x\Big(f(x_0)+f(x_1)\Big)+\frac{1}{2}\Delta x\Big(f(x_1)+f(x_2)\Big)+\frac{1}{2}\Delta x\Big(f(x_2)+f(x_3)\Big)+\frac{1}{2}\Delta x\Big(f(x_3)+f(x_4)\Big).$$

After taking out a common factor of $\frac{1}{2}\Delta x$ and combining like terms, we have

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{2}\Big[f(x_0)+2\,f(x_1)+2\,f(x_2)+2\,f(x_3)+f(x_4)\Big].$$

Generalizing, we formally state the following rule.

The Trapezoidal Rule

Assume that $f(x)$ is continuous over $[a,b]$. Let n be a positive integer and $\Delta x = \dfrac{b-a}{n}$. Let $[a,b]$ be divided into n subintervals, each of length Δx, with endpoints at $P=\{x_0,x_1,x_2\ldots,x_n\}$.

Set

$$T_n = \frac{\Delta x}{2}\Big[f(x_0)+2\,f(x_1)+2\,f(x_2)+\cdots+2\,f(x_{n-1})+f(x_n)\Big].$$

Then, $\lim\limits_{n\to+\infty} T_n = \int_a^b f(x)\,dx$.

Before continuing, let's make a few observations about the trapezoidal rule. First of all, it is useful to note that

$$T_n = \frac{1}{2}(L_n+R_n) \text{ where } L_n = \sum_{i=1}^{n} f(x_{i-1})\Delta x \text{ and } R_n = \sum_{i=1}^{n} f(x_i)\Delta x.$$

That is, L_n and R_n approximate the integral using the left-hand and right-hand endpoints of each subinterval, respectively. In addition, a careful examination of Figure 3.6.3 leads us to make the following observations about using the trapezoidal rules and midpoint rules to estimate the definite integral of a nonnegative function. The trapezoidal rule tends to overestimate the value of a definite integral systematically over intervals where the function is concave up and to underestimate the value of a definite integral systematically over intervals where the function is concave down. On the other hand, the midpoint rule tends to average out these errors somewhat by partially overestimating and partially underestimating the value of the definite integral over these same types of intervals. This leads us to hypothesize that, in general, the midpoint rule tends to be more accurate than the trapezoidal rule.

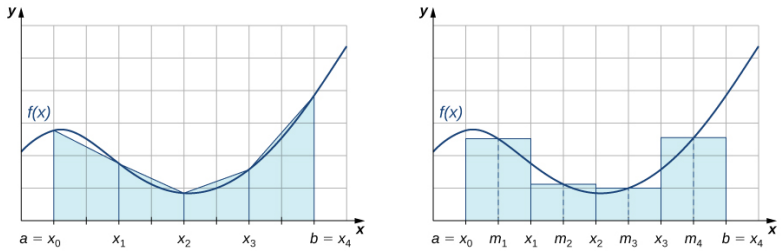


Figure 3.6.3:The trapezoidal rule tends to be less accurate than the midpoint rule.

Example 3.6.3: Using the Trapezoidal Rule

Use the trapezoidal rule to estimate $\int_0^1 x^2\,dx$ using four subintervals.

Solution

The endpoints of the subintervals consist of elements of the set $P=\left\{0,\frac{1}{4},\frac{1}{2},\frac{3}{4},1\right\}$ and $\Delta x = \frac{1-0}{4}=\frac{1}{4}$. Thus,

$$\begin{aligned}\int_0^1 x^2dx &\approx \frac{1}{2}\cdot\frac{1}{4}\Big[f(0)+2\,f\left(\tfrac{1}{4}\right)+2\,f\left(\tfrac{1}{2}\right)+2\,f\left(\tfrac{3}{4}\right)+f(1)\Big]\\ &= \tfrac{1}{8}\left(0+2\cdot\tfrac{1}{16}+2\cdot\tfrac{1}{4}+2\cdot\tfrac{9}{16}+1\right)\\ &= \frac{11}{32}=0.34375\end{aligned}$$

Exercise 3.6.2

Use the trapezoidal rule with $n=2$ to estimate $\int_1^2 \frac{1}{x}\,dx$.

Hint

Set $\Delta x = \dfrac{1}{2}$. The endpoints of the subintervals are the elements of the set $P=\left\{1,\frac{3}{2},2\right\}$.

Answer

$\frac{17}{24}\approx 0.708333$

Absolute and Relative Error

An important aspect of using these numerical approximation rules consists of calculating the error in using them for estimating the value of a definite integral. We first need to define absolute error and relative error.

Definition: absolute and relative error

If B is our estimate of some quantity having an actual value of A, then the **absolute error** is given by $|A-B|$.

The **relative error** is the error as a percentage of the actual value and is given by

$$\left|\frac{A-B}{A}\right| \cdot 100\%.$$

Example 3.6.4: Calculating Error in the Midpoint Rule

Calculate the absolute and relative error in the estimate of $\int_0^1 x^2\,dx$ using the midpoint rule, found in Example 3.6.1.

Solution: The calculated value is $\int_0^1 x^2\,dx = \frac{1}{3}$ and our estimate from the example is $M_4 = \frac{21}{64}$. Thus, the absolute error is given by $\left|\frac{1}{3} - \frac{21}{64}\right| = \frac{1}{192} \approx 0.0052$.

The relative error is

$$\frac{1/192}{1/3} = \frac{1}{64} \approx 0.015625 \approx 1.6\%.$$

Example 3.6.5: Calculating Error in the Trapezoidal Rule

Calculate the absolute and relative error in the estimate of $\int_0^1 x^2\,dx$ using the trapezoidal rule, found in Example 3.6.3.

Solution: The calculated value is $\int_0^1 x^2\,dx = \frac{1}{3}$ and our estimate from the example is $T_4 = \frac{11}{32}$. Thus, the absolute error is given by $\left|\frac{1}{3} - \frac{11}{32}\right| = \frac{1}{96} \approx 0.0104$.

The relative error is given by

$$\frac{1/96}{1/3} = 0.03125 \approx 3.1\%.$$

Exercise 3.6.3

In an earlier checkpoint, we estimated $\int_1^2 \frac{1}{x}\,dx$ to be $\frac{24}{35}$ using M_2. The actual value of this integral is $\ln 2$. Using $\frac{24}{35} \approx 0.6857$ and $\ln 2 \approx 0.6931$, calculate the absolute error and the relative error.

Hint

Use the previous examples as a guide.

Answer

absolute error ≈ 0.0074,and relative error $\approx 1.1\%$

Error Bounds on the Midpoint and Trapezoidal Rules

In the two previous examples, we were able to compare our estimate of an integral with the actual value of the integral; however, we do not typically have this luxury. In general, if we are approximating an integral, we are doing so because we cannot compute the exact value of the integral itself easily. Therefore, it is often helpful to be able to determine an upper bound for the error in an approximation of an integral. The following theorem provides error bounds for the midpoint and trapezoidal rules. The theorem is stated without proof.

Error Bounds for the Midpoint and Trapezoidal Rules

Let $f(x)$ be a continuous function over $[a,b]$, having a second derivative $f''(x)$ over this interval. If M is the maximum value of $|f''(x)|$ over $[a,b]$, then the upper bounds for the error in using M_n and T_n to estimate $\int_a^b f(x)\,dx$ are

$$\text{Error in } M_n \le \frac{M(b-a)^3}{24n^2} \tag{3.6.1}$$

and

$$\text{Error in } T_n \le \frac{M(b-a)^3}{12n^2}$$

.

We can use these bounds to determine the value of n necessary to guarantee that the error in an estimate is less than a specified value.

Example 3.6.6: Determining the Number of Intervals to Use

What value of n should be used to guarantee that an estimate of $\int_0^1 e^{x^2}\,dx$ is accurate to within 0.01 if we use the midpoint rule?

Solution

We begin by determining the value of M, the maximum value of $|f''(x)|$ over $[0,1]$ for $f(x) = e^{x^2}$. Since $f'(x) = 2xe^{x^2}$, we have

$$f''(x) = 2e^{x^2} + 4x^2e^{x^2}.$$

Thus,

$$|f''(x)| = 2e^{x^2}(1+2x^2) \le 2 \cdot e \cdot 3 = 6e.$$

From the error-bound Equation 3.6.1, we have

$$\text{Error in } M_n \le \frac{M(b-a)^3}{24n^2} \le \frac{6e(1-0)^3}{24n^2} = \frac{6e}{24n^2}.$$

Now we solve the following inequality for n:

$$\frac{6e}{24n^2} \le 0.01.$$

Thus, $n \ge \sqrt{\frac{600e}{24}} \approx 8.24$. Since n must be an integer satisfying this inequality, a choice of $n = 9$ would guarantee that

$$\left| \int_0^1 e^{x^2}\,dx - M_n \right| < 0.01.$$

Analysis

We might have been tempted to round 8.24 down and choose $n = 8$, but this would be incorrect because we must have an integer greater than or equal to 8.24. We need to keep in mind that the error estimates provide an upper bound only for the error. The actual estimate may, in fact, be a much better approximation than is indicated by the error bound.

? Exercise 3.6.4

Use Equation 3.6.1 to find an upper bound for the error in using M_4 to estimate $\int_0^1 x^2\,dx$.

Hint

$f''(x) = 2,$ so $M = 2.$

Answer

$\frac{1}{192}$

Simpson's Rule

With the midpoint rule, we estimated areas of regions under curves by using rectangles. In a sense, we approximated the curve with piecewise constant functions. With the trapezoidal rule, we approximated the curve by using piecewise linear functions. What if we were, instead, to approximate a curve using piecewise quadratic functions? With **Simpson's rule,** we do just this. We partition the interval into an even number of subintervals, each of equal width. Over the first pair of subintervals we approximate $\int_{x_0}^{x_2} f(x)\,dx$ with $\int_{x_0}^{x_2} p(x)\,dx$, where $p(x) = Ax^2 + Bx + C$ is the quadratic function passing through $(x_0, f(x_0))$, $(x_1, f(x_1))$, and $(x_2, f(x_2))$ (Figure 3.6.4). Over the next pair of subintervals we approximate $\int_{x_2}^{x_4} f(x)\,dx$ with the integral of another quadratic function passing through $(x_2, f(x_2))$, $(x_3, f(x_3))$, and $(x_4, f(x_4))$. This process is continued with each successive pair of subintervals.

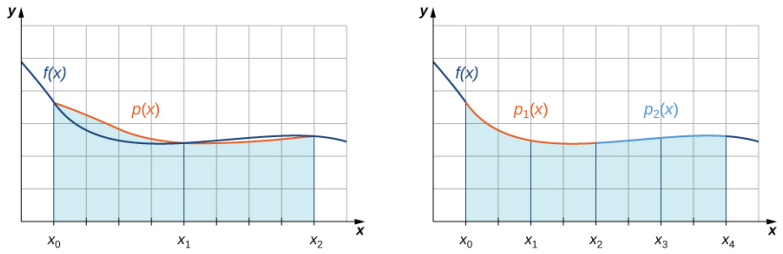


Figure 3.6.4: With Simpson's rule, we approximate a definite integral by integrating a piecewise quadratic function.

To understand the formula that we obtain for Simpson's rule, we begin by deriving a formula for this approximation over the first two subintervals. As we go through the derivation, we need to keep in mind the following relationships:

$$f(x_0) = p(x_0) = Ax_0^2 + Bx_0 + C$$
$$f(x_1) = p(x_1) = Ax_1^2 + Bx_1 + C$$
$$f(x_2) = p(x_2) = Ax_2^2 + Bx_2 + C$$

$x_2 - x_0 = 2\Delta x$, where Δx is the length of a subinterval.

$$x_2 + x_0 = 2x_1, \text{ since } x_1 = \frac{(x_2 + x_0)}{2}.$$

Thus,

$$\int_{x_0}^{x_2} f(x)\,dx \approx \int_{x_0}^{x_2} p(x)\,dx$$

$$= \int_{x_0}^{x_2} (Ax^2+Bx+C)\,dx$$

$$= \left(\frac{A}{3}x^3+\frac{B}{2}x^2+Cx\right)\Big|_{x_0}^{x_2} \qquad \text{Find the antiderivative.}$$

$$= \frac{A}{3}(x_2^3-x_0^3)+\frac{B}{2}(x_2^2-x_0^2)+C(x_2-x_0) \qquad \text{Evaluate the antiderivative.}$$

$$= \frac{A}{3}(x_2-x_0)(x_2^2+x_2x_0+x_0^2)+\frac{B}{2}(x_2-x_0)(x_2+x_0)+C(x_2-x_0)$$

$$= \frac{x_2-x_0}{6}\left(2A(x_2^2+x_2x_0+x_0^2)+3B(x_2+x_0)+6C\right) \qquad \text{Factor out } \frac{x_2-x_0}{6}.$$

$$= \frac{\Delta x}{3}\left((Ax_2^2+Bx_2+C)+(Ax_0^2+Bx_0+C)+A(x_2^2+2x_2x_0+x_0^2)+2B(x_2+x_0)+4C\right) \qquad \text{Rearrange the terms. Note: } \Delta x = \frac{x_2-x_0}{2}$$

$$= \frac{\Delta x}{3}\left(f(x_2)+f(x_0)+A(x_2+x_0)^2+2B(x_2+x_0)+4C\right) \qquad \text{Factor and substitute:}$$

$$f(x_2) = Ax_2^2+Bx_2+C \text{ and } f(x_0) = Ax_0^2+Bx_0+C.$$

$$= \frac{\Delta x}{3}\left(f(x_2)+f(x_0)+A(2x_1)^2+2B(2x_1)+4C\right) \qquad \text{Substitute } x_2+x_0 = 2x_1.$$

Note: $x_1 = \dfrac{x_2+x_0}{2}$ is the midpoint.

$$= \frac{\Delta x}{3}\left(f(x_2)+4f(x_1)+f(x_0)\right). \qquad \text{Expand and substitute } f(x_1) = Ax_1^2+Bx_1+C.$$

If we approximate $\int_{x_2}^{x_4} f(x)\,dx$ using the same method, we see that we have

$$\int_{x_2}^{x_4} f(x)\,dx \approx \frac{\Delta x}{3}(f(x_4)+4\,f(x_3)+f(x_2)).$$

Combining these two approximations, we get

$$\int_{x_0}^{x_4} f(x)\,dx \approx \frac{\Delta x}{3}(f(x_0)+4\,f(x_1)+2\,f(x_2)+4\,f(x_3)+f(x_4)).$$

The pattern continues as we add pairs of subintervals to our approximation. The general rule may be stated as follows.

Simpson's Rule

Assume that $f(x)$ is continuous over $[a,b]$. Let n be a positive even integer and $\Delta x = \dfrac{b-a}{n}$. Let $[a,b]$ be divided into n subintervals, each of length Δx, with endpoints at $P = \{x_0, x_1, x_2, \ldots, x_n\}$. Set

$$S_n = \frac{\Delta x}{3}\left[f(x_0)+4\,f(x_1)+2\,f(x_2)+4\,f(x_3)+2\,f(x_4)+\cdots+2\,f(x_{n-2})+4\,f(x_{n-1})+f(x_n)\right].$$

Then,

$$\lim_{n\to+\infty} S_n = \int_a^b f(x)\,dx.$$

Just as the trapezoidal rule is the average of the left-hand and right-hand rules for estimating definite integrals, Simpson's rule may be obtained from the midpoint and trapezoidal rules by using a weighted average. It can be shown that $S_{2n} = \left(\frac{2}{3}\right) M_n + \left(\frac{1}{3}\right) T_n$.

It is also possible to put a bound on the error when using Simpson's rule to approximate a definite integral. The bound in the error is given by the following rule:

Rule: Error Bound for Simpson's Rule

Let $f(x)$ be a continuous function over $[a,b]$ having a fourth derivative, $f^{(4)}(x)$, over this interval. If M is the maximum value of $\left|f^{(4)}(x)\right|$ over $[a,b]$, then the upper bound for the error in using S_n to estimate $\int_a^b f(x)\,dx$ is given by

$$\text{Error in } S_n \le \frac{M(b-a)^5}{180n^4}.$$

Example 3.6.7: Applying Simpson's Rule 1

Use S_2 to approximate $\int_0^1 x^3\,dx$. Estimate a bound for the error in S_2.

Solution

Since $[0,1]$ is divided into two intervals, each subinterval has length $\Delta x = \frac{1-0}{2} = \frac{1}{2}$. The endpoints of these subintervals are $\left\{0, \frac{1}{2}, 1\right\}$. If we set $f(x) = x^3$, then

$$S_2 = \frac{1}{3}\cdot\frac{1}{2}(f(0)+4\,f(\tfrac{1}{2})+f(1)) = \frac{1}{6}(0+4\cdot\frac{1}{8}+1) = \frac{1}{4}.$$

Since $f^{(4)}(x) = 0$ and consequently $M = 0$, we see that

$$\text{Error in } S_2 \le \frac{0(1)^5}{180 \cdot 2^4} = 0.$$

This bound indicates that the value obtained through Simpson's rule is exact. A quick check will verify that, in fact, $\int_0^1 x^3\,dx = \frac{1}{4}$.

Example 3.6.8: Applying Simpson's Rule 2

Use S_6 to estimate the length of the curve $y = \frac{1}{2}x^2$ over $[1, 4]$.

Solution

The length of $y = \frac{1}{2}x^2$ over $[1, 4]$ is $\int_1^4 \sqrt{1+x^2}\,dx$. If we divide $[1, 4]$ into six subintervals, then each subinterval has length $\Delta x = \frac{4-1}{6} = \frac{1}{2}$, and the endpoints of the subintervals are $\left\{1, \frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}, 4\right\}$. Setting $f(x) = \sqrt{1+x^2}$,

$$S_6 = \frac{1}{3} \cdot \frac{1}{2}(f(1) + 4f(\frac{3}{2}) + 2f(2) + 4f(\frac{5}{2}) + 2f(3) + 4f(\frac{7}{2}) + f(4)).$$

After substituting, we have

$$S_6 = \frac{1}{6}(1.4142 + 4 \cdot 1.80278 + 2 \cdot 2.23607 + 4 \cdot 2.69258 + 2 \cdot 3.16228 + 4 \cdot 3.64005 + 4.12311) \approx 8.14594 \text{ units}.$$

Exercise 3.6.5

Use S_2 to estimate $\int_1^2 \frac{1}{x}\,dx$.

Hint

$$S_2 = \frac{1}{3}\Delta x\,(f(x_0) + 4f(x_1) + f(x_2))$$

Answer

$\frac{25}{36} \approx 0.694444$

Key Concepts

- We can use numerical integration to estimate the values of definite integrals when a closed form of the integral is difficult to find or when an approximate value only of the definite integral is needed.
- The most commonly used techniques for numerical integration are the midpoint rule, trapezoidal rule, and Simpson's rule.
- The midpoint rule approximates the definite integral using rectangular regions whereas the trapezoidal rule approximates the definite integral using trapezoidal approximations.
- Simpson's rule approximates the definite integral by first approximating the original function using piecewise quadratic functions.

Key Equations

- **Midpoint rule**

  $M_n = \sum_{i=1}^{n} f(m_i)\Delta x$

- **Trapezoidal rule**

  $T_n = \frac{\Delta x}{2}\left[f(x_0) + 2\,f(x_1) + 2\,f(x_2) + \cdots + 2\,f(x_{n-1}) + f(x_n)\right]$

- **Simpson's rule**

  $S_n = \frac{\Delta x}{3}\left[f(x_0) + 4\,f(x_1) + 2\,f(x_2) + 4\,f(x_3) + 2\,f(x_4) + 4\,f(x_5) + \cdots + 2\,f(x_{n-2}) + 4\,f(x_{n-1}) + f(x_n)\right]$

- **Error bound for midpoint rule**

  Error in $M_n \le \frac{M(b-a)^3}{24n^2}$, where M is the maximum value of $|f''(x)|$ over $[a, b]$.

- **Error bound for trapezoidal rule**

  Error in $T_n \le \frac{M(b-a)^3}{12n^2}$, where M is the maximum value of $|f''(x)|$ over $[a, b]$.

- **Error bound for Simpson's rule**

  Error in $S_n \le \frac{M(b-a)^5}{180n^4}$, where M is the maximum value of $\left|f^{(4)}(x)\right|$ over $[a, b]$.

Glossary

absolute error

if B is an estimate of some quantity having an actual value of A, then the absolute error is given by $|A - B|$

midpoint rule

a rule that uses a Riemann sum of the form $M_n = \sum_{i=1}^{n} f(m_i)\Delta x$, where m_i is the midpoint of the ith subinterval to approximate $\int_a^b f(x)\,dx$

numerical integration

the variety of numerical methods used to estimate the value of a definite integral, including the midpoint rule, trapezoidal rule, and Simpson's rule

relative error

error as a percentage of the actual value, given by

$$\text{relative error} = \left| \frac{A-B}{A} \right| \cdot 100\%$$

Simpson's rule

a rule that approximates $\int_a^b f(x)\,dx$ using the area under a piecewise quadratic function.

The approximation S_n to $\int_a^b f(x)\,dx$ is given by

$$S_n = \frac{\Delta x}{3}\big(f(x_0) + 4\,f(x_1) + 2\,f(x_2) + 4\,f(x_3) + 2\,f(x_4) + \cdots + 2\,f(x_{n-2}) + 4\,f(x_{n-1}) + f(x_n)\big).$$

trapezoidal rule

a rule that approximates $\int_a^b f(x)\,dx$ using the area of trapezoids.

The approximation T_n to $\int_a^b f(x)\,dx$ is given by

$$T_n = \frac{\Delta x}{2}\big(f(x_0) + 2\,f(x_1) + 2\,f(x_2) + \cdots + 2\,f(x_{n-1}) + f(x_n)\big).$$

- Gilbert Strang (MIT) and Edwin "Jed" Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.
- Edited by Paul Seeburger (Monroe Community College). Notes added to development of area under a parabola and typos fixed in original text.

This page titled 3.6: Numerical Integration is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.6: Numerical Integration** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

3.7: Improper Integrals

Learning Objectives

- Evaluate an integral over an infinite interval.
- Evaluate an integral over a closed interval with an infinite discontinuity within the interval.
- Use the comparison theorem to determine whether a definite integral is convergent.

Is the area between the graph of $f(x) = \dfrac{1}{x}$ and the x-axis over the interval $[1, +\infty)$ finite or infinite? If this same region is revolved about the x-axis, is the volume finite or infinite? Surprisingly, the area of the region described is infinite, but the volume of the solid obtained by revolving this region about the x-axis is finite.

In this section, we define integrals over an infinite interval as well as integrals of functions containing a discontinuity on the interval. Integrals of these types are called improper integrals. We examine several techniques for evaluating improper integrals, all of which involve taking limits.

Integrating over an Infinite Interval

How should we go about defining an integral of the type $\displaystyle\int_a^{+\infty} f(x)\,dx$? We can integrate $\displaystyle\int_a^t f(x)\,dx$ for any value of t, so it is reasonable to look at the behavior of this integral as we substitute larger values of t. Figure 3.7.1 shows that $\displaystyle\int_a^t f(x)\,dx$ may be interpreted as area for various values of t. In other words, we may define an improper integral as a limit, taken as one of the limits of integration increases or decreases without bound.

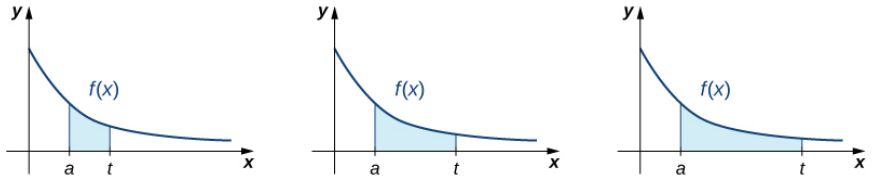


Figure 3.7.1: To integrate a function over an infinite interval, we consider the limit of the integral as the upper limit increases without bound.

Definition: Improper Integral

1. Let $f(x)$ be continuous over an interval of the form $[a, +\infty)$. Then

$$\int_a^{+\infty} f(x)\,dx = \lim_{t\to+\infty} \int_a^t f(x)\,dx, \tag{3.7.1}$$

   provided this limit exists.

2. Let $f(x)$ be continuous over an interval of the form $(-\infty, b]$. Then

$$\int_{-\infty}^b f(x)\,dx = \lim_{t\to-\infty} \int_t^b f(x)\,dx, \tag{3.7.2}$$

   provided this limit exists.

In each case, if the limit exists, then the *improper integral* is said to converge. If the limit does not exist, then the improper integral is said to diverge.

3. Let $f(x)$ be continuous over $(-\infty, +\infty)$. Then

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^0 f(x)\,dx + \int_0^{+\infty} f(x)\,dx, \tag{3.7.3}$$

   provided that $\displaystyle\int_{-\infty}^0 f(x)\,dx$ and $\displaystyle\int_0^{+\infty} f(x)\,dx$ both converge.

   If either of these two integrals diverge, then $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx$ diverges.

(It can be shown that, in fact, $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^a f(x)\,dx + \int_a^{+\infty} f(x)\,dx$ for any value of a.).

In our first example, we return to the question we posed at the start of this section: Is the area between the graph of $f(x) = \dfrac{1}{x}$ and the x-axis over the interval $[1, +\infty)$ finite or infinite?

Example 3.7.1: Finding an Area

Determine whether the area between the graph of $f(x) = \dfrac{1}{x}$ and the x-axis over the interval $[1, +\infty)$ is finite or infinite.

Solution

We first do a quick sketch of the region in question, as shown in Figure 3.7.2.

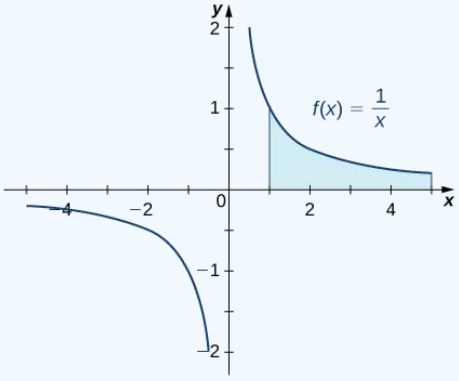


Figure 3.7.2: We can find the area between the curve $f(x) = 1/x$ and the x-axis on an infinite interval.

We can see that the area of this region is given by

$$A = \int_1^{\infty} \frac{1}{x}\, dx.$$

which can be evaluated using Equation 3.7.1:

$$\begin{aligned} A &= \int_1^{\infty} \frac{1}{x}\, dx \\ &= \lim_{t\to+\infty} \int_1^{t} \frac{1}{x}\, dx && \text{Rewrite the improper integral as a limit} \\ &= \lim_{t\to+\infty} \ln|x|\Big|_1^{t} && \text{Find the antiderivative} \\ &= \lim_{t\to+\infty} (\ln|t| - \ln 1) && \text{Evaluate the antiderivative} \\ &= +\infty. && \text{Evaluate the limit.} \end{aligned}$$

Since the improper integral diverges to $+\infty$, the area of the region is infinite.

Example 3.7.2: Finding a Volume

Find the volume of the solid obtained by revolving the region bounded by the graph of $f(x) = \dfrac{1}{x}$ and the x-axis over the interval $[1, +\infty)$ about the x-axis.

Solution

The solid is shown in Figure 3.7.3. Using the disk method, we see that the volume V is

$$V = \pi \int_1^{+\infty} \frac{1}{x^2}\, dx.$$

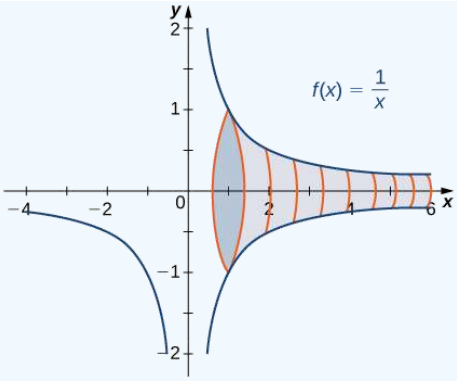


Figure 3.7.3: The solid of revolution can be generated by rotating an infinite area about the x-axis.

Then we have

$$
\begin{aligned}
V &= \pi \int_1^{+\infty} \frac{1}{x^2}\,dx \\
&= \pi \lim_{t\to+\infty} \int_1^t \frac{1}{x^2}\,dx && \text{Rewrite as a limit.} \\
&= \pi \lim_{t\to+\infty} -\frac{1}{x}\Big|_1^t && \text{Find the antiderivative.} \\
&= \pi \lim_{t\to+\infty} \left(-\frac{1}{t}+1\right) && \text{Evaluate the antiderivative.} \\
&= \pi \lim_{t\to+\infty} \left(-\frac{1}{\cancel{t}^{\,0}}+1\right) && \text{Evaluate the antiderivative.} \\
&= \pi
\end{aligned}
$$

The improper integral converges to π. Therefore, the volume of the solid of revolution is π.

In conclusion, although the area of the region between the x-axis and the graph of $f(x) = 1/x$ over the interval $[1, +\infty)$ is infinite, the volume of the solid generated by revolving this region about the x-axis is finite. The solid generated is known as Gabriel's Horn.

Note: Gabriel's horn (also called Torricelli's trumpet) is a geometric figure which has *infinite* surface area, but *finite* volume. The name refers to the tradition identifying the Archangel Gabriel as the angel who blows the horn to announce Judgment Day, associating the divine, or infinite, with the finite. The properties of this figure were first studied by Italian physicist and mathematician Evangelista Torricelli in the 17th century.

Example 3.7.3: Traffic Accidents in a City

Suppose that at a busy intersection, traffic accidents occur at an average rate of one every three months. After residents complained, changes were made to the traffic lights at the intersection. It has now been eight months since the changes were made and there have been no accidents. Were the changes effective or is the 8-month interval without an accident a result of chance?



Figure 3.7.4: Modification of work by David McKelvey, Flickr.

Probability theory tells us that if the average time between events is k, the *probability* that X, the time between events, is between a and b is given by

$$(P(a \le x \le b) = \int_a^b f(x)\,dx$$

where

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ ke^{-kx}, & \text{if } x \geq 0 \end{cases}.$$

Thus, if accidents are occurring at a rate of one every 3 months, then the probability that X, the time between accidents, is between a and b is given by

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

where

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ 3e^{-3x}, & \text{if } x \geq 0 \end{cases}.$$

To answer the question, we must compute $P(X \geq 8) = \int_8^{+\infty} 3e^{-3x}\,dx$ and decide whether it is likely that 8 months could have passed without an accident if there had been no improvement in the traffic situation.

Solution

We need to calculate the probability as an improper integral:

$$\begin{aligned} P(X \geq 8) &= \int_8^{+\infty} 3e^{-3x}\,dx \\ &= \lim_{t \to +\infty} \int_8^t 3e^{-3x}\,dx \\ &= \lim_{t \to +\infty} -e^{-3x}\Big|_8^t \\ &= \lim_{t \to +\infty} (-e^{-3t} + e^{-24}) \\ &\approx 3.8 \times 10^{-11}. \end{aligned}$$

The value 3.8×10^{-11} represents the probability of no accidents in 8 months under the initial conditions. Since this value is very, very small, it is reasonable to conclude the changes were effective.

Example 3.7.4: Evaluating an Improper Integral over an Infinite Interval

Evaluate $\int_{-\infty}^0 \frac{1}{x^2+4}\,dx$. State whether the improper integral converges or diverges.

Solution

Begin by rewriting $\int_{-\infty}^0 \frac{1}{x^2+4}\,dx$ as a limit using Equation 3.7.2 from the definition. Thus,

$$\begin{aligned} \int_{-\infty}^0 \frac{1}{x^2+4}\,dx &= \lim_{t \to -\infty} \int_t^0 \frac{1}{x^2+4}\,dx && \text{Rewrite as a limit.} \\ &= \lim_{t \to -\infty} \frac{1}{2}\tan^{-1}\frac{x}{2}\Big|_t^0 && \text{Find the antiderivative.} \\ &= \lim_{t \to -\infty} \left(\frac{1}{2}\tan^{-1} 0 - \frac{1}{2}\tan^{-1}\frac{t}{2}\right) && \text{Evaluate the antiderivative.} \\ &= \frac{\pi}{4}. && \text{Evaluate the limit and simplify.} \end{aligned}$$

The improper integral converges to $\frac{\pi}{4}$.

Example 3.7.5: Evaluating an Improper Integral on $(-\infty, +\infty)$

Evaluate $\int_{-\infty}^{+\infty} xe^x\,dx$. State whether the improper integral converges or diverges.

Solution

Start by splitting up the integral:

$$\int_{-\infty}^{+\infty} xe^x\,dx = \int_{-\infty}^{0} xe^x\,dx + \int_{0}^{+\infty} xe^x\,dx.$$

If either $\int_{-\infty}^{0} xe^x\,dx$ or $\int_{0}^{+\infty} xe^x\,dx$ diverges, then $\int_{-\infty}^{+\infty} xe^x\,dx$ diverges.

Compute each integral separately.

For the first integral,

$$\begin{aligned}
\int_{-\infty}^{0} xe^x\,dx &= \lim_{t\to-\infty}\int_{t}^{0} xe^x\,dx && \text{Rewrite as a limit.}\\
&= \lim_{t\to-\infty}(xe^x - e^x)\Big|_{t}^{0} && \text{Use integration by parts to find the antiderivative. (Here } u = x \text{ and } dv = e^x\,dx.\text{)}\\
&= \lim_{t\to-\infty}(-1 - te^t + e^t) && \text{Evaluate the antiderivative.}\\
&= -1.
\end{aligned}$$

Evaluate the limit. Note: $\lim_{t\to-\infty} te^t$ is indeterminate of the form $0\cdot\infty$.Thus, $\lim_{t\to-\infty} te^t = \lim_{t\to-\infty}\frac{t}{e^{-t}} = \lim_{t\to-\infty}\frac{-1}{e^{-t}} = \lim_{t\to-\infty} -e^t = 0$ by L'Hôpital's Rule.

The first improper integral converges. For the second integral,

$$\begin{aligned}
\int_{0}^{+\infty} xe^x\,dx &= \lim_{t\to+\infty}\int_{0}^{t} xe^x\,dx && \text{Rewrite as a limit.}\\
&= \lim_{t\to+\infty}(xe^x - e^x)\Big|_{0}^{t} && \text{Find the antiderivative.}\\
&= \lim_{t\to+\infty}(te^t - e^t + 1) && \text{Evaluate the antiderivative.}\\
&= \lim_{t\to+\infty}((t-1)e^t + 1) && \text{Rewrite. } (te^t - e^t \text{ is indeterminate.})\\
&= +\infty. && \text{Evaluate the limit.}
\end{aligned}$$

Thus, $\int_{0}^{+\infty} xe^x\,dx$ diverges. Since this integral diverges, $\int_{-\infty}^{+\infty} xe^x\,dx$ diverges as well.

? Exercise 3.7.1

Evaluate $\int_{-3}^{+\infty} e^{-x}\,dx$. State whether the improper integral converges or diverges.

Hint

$$\int_{-3}^{+\infty} e^{-x}\,dx = \lim_{t\to+\infty}\int_{-3}^{t} e^{-x}\,dx$$

Answer

It converges to e^3.

Integrating a Discontinuous Integrand

Now let's examine integrals of functions containing an infinite discontinuity in the interval over which the integration occurs. Consider an integral of the form $\int_a^b f(x)\,dx$, where $f(x)$ is continuous over $[a,b)$ and discontinuous at b. Since the function $f(x)$ is continuous over $[a,t]$ for all values of t satisfying $a \le t < b$, the integral $\int_a^t f(x)\,dx$ is defined for all such values of t. Thus, it makes sense to consider the values of $\int_a^t f(x)\,dx$ as t approaches b for $a \le t < b$. That is, we define $\int_a^b f(x)\,dx = \lim_{t\to b^-}\int_a^t f(x)\,dx$, provided this limit exists. Figure 3.7.5 illustrates $\int_a^t f(x)\,dx$ as areas of regions for values of t approaching b.

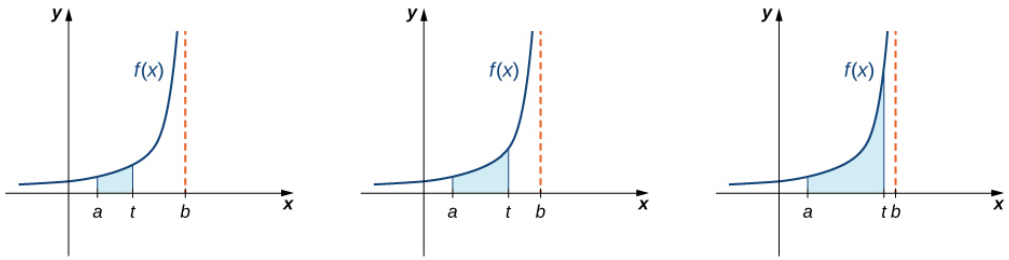


Figure 3.7.5: As t approaches b from the left, the value of the area from a to t approaches the area from a to b.

We use a similar approach to define $\int_a^b f(x)\,dx$, where $f(x)$ is continuous over $(a, b]$ and discontinuous at a. We now proceed with a formal definition.

Definition: Converging and Diverging Improper Integral

1. Let $f(x)$ be continuous over $[a, b)$. Then,

$$\int_a^b f(x)\,dx = \lim_{t\to b^-} \int_a^t f(x)\,dx. \tag{3.7.4}$$

2. Let $f(x)$ be continuous over $(a, b]$. Then,

$$\int_a^b f(x)\,dx = \lim_{t\to a^+} \int_t^b f(x)\,dx. \tag{3.7.5}$$

In each case, if the limit exists, then the improper integral is said to converge. If the limit does not exist, then the improper integral is said to diverge.

3. If $f(x)$ is continuous over $[a, b]$ except at a point c in (a, b), then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx, \tag{3.7.6}$$

provided both $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ converge. If either of these integrals diverges, then $\int_a^b f(x)\,dx$ diverges.

The following examples demonstrate the application of this definition.

Example 3.7.6: Integrating a Discontinuous Integrand

Evaluate $\int_0^4 \frac{1}{\sqrt{4-x}}\,dx$, if possible. State whether the integral converges or diverges.

Solution

The function $f(x) = \frac{1}{\sqrt{4-x}}$ is continuous over $[0, 4)$ and discontinuous at 4. Using Equation 3.7.4 from the definition, rewrite $\int_0^4 \frac{1}{\sqrt{4-x}}\,dx$ as a limit:

$$\begin{aligned}
\int_0^4 \frac{1}{\sqrt{4-x}}\,dx &= \lim_{t\to 4^-} \int_0^t \frac{1}{\sqrt{4-x}}\,dx && \text{Rewrite as a limit.}\\
&= \lim_{t\to 4^-} \left(-2\sqrt{4-x}\right)\Big|_0^t && \text{Find the antiderivative.}\\
&= \lim_{t\to 4^-} \left(-2\sqrt{4-t}+4\right) && \text{Evaluate the antiderivative.}\\
&= 4. && \text{Evaluate the limit.}
\end{aligned}$$

The improper integral converges.

Example 3.7.7: Integrating a Discontinuous Integrand

Evaluate $\int_0^2 x \ln x \, dx$. State whether the integral converges or diverges.

Solution

Since $f(x) = x \ln x$ is continuous over $(0, 2]$ and is discontinuous at zero, we can rewrite the integral in limit form using Equation 3.7.5:

$$\int_0^2 x \ln x \, dx = \lim_{t \to 0^+} \int_t^2 x \ln x \, dx \qquad \text{Rewrite as a limit.}$$

$$= \lim_{t \to 0^+} \left(\frac{1}{2} x^2 \ln x - \frac{1}{4} x^2 \right) \Big|_t^2 \qquad \text{Evaluate } \int x \ln x \, dx \text{ using integration by parts with } u = \ln x \text{ and } dv = x.$$

$$= \lim_{t \to 0^+} \left(2 \ln 2 - 1 - \frac{1}{2} t^2 \ln t + \frac{1}{4} t^2 \right). \qquad \text{Evaluate the antiderivative.}$$

$$= 2 \ln 2 - 1. \qquad \text{Evaluate the limit. Note that } \lim_{t \to 0^+} t^2 \ln t \text{ is indeterminate.}$$

To evaluate it, rewrite it as a quotient and apply L'Hôpital's rule.

The improper integral converges.

Example 3.7.8: Integrating a Discontinuous Integrand

Evaluate $\int_{-1}^1 \frac{1}{x^3} \, dx$. State whether the improper integral converges or diverges.

Solution

Since $f(x) = 1/x^3$ is discontinuous at zero, using Equation 3.7.6, we can write

$$\int_{-1}^1 \frac{1}{x^3} \, dx = \int_{-1}^0 \frac{1}{x^3} \, dx \quad + \quad \int_0^1 \frac{1}{x^3} \, dx.$$

If either of the two integrals diverges, then the original integral diverges. Begin with $\int_{-1}^0 \frac{1}{x^3} \, dx$:

$$\int_{-1}^0 \frac{1}{x^3} \, dx = \lim_{t \to 0^-} \int_{-1}^t \frac{1}{x^3} \, dx \qquad \text{Rewrite as a limit.}$$

$$= \lim_{t \to 0^-} \left(-\frac{1}{2x^2} \right) \Big|_{-1}^t \qquad \text{Find the antiderivative.}$$

$$= \lim_{t \to 0^-} \left(-\frac{1}{2t^2} + \frac{1}{2} \right) \qquad \text{Evaluate the antiderivative.}$$

$$= +\infty. \qquad \text{Evaluate the limit.}$$

Therefore, $\int_{-1}^0 \frac{1}{x^3} \, dx$ diverges.

Since $\int_{-1}^0 \frac{1}{x^3} \, dx$ diverges, $\int_{-1}^1 \frac{1}{x^3} \, dx$ also diverges.

Exercise 3.7.2

Evaluate $\int_0^2 \frac{1}{x} \, dx$. State whether the integral converges or diverges.

Hint

Write $\int_0^2 \frac{1}{x} \, dx$ in limit form using Equation 3.7.5

Answer

$+\infty$, It diverges.

A Comparison Theorem

It is not always easy or even possible to evaluate an improper integral directly; however, by comparing it with another carefully chosen integral, it may be possible to determine its convergence or divergence. To see this, consider two continuous functions $f(x)$ and $g(x)$ satisfying $0 \le f(x) \le g(x)$ for $x \ge a$ (Figure 3.7.6). In this case, we may view integrals of these functions over intervals of the form $[a, t]$ as areas, so we have the relationship

$$0 \le \int_a^t f(x)\,dx \le \int_a^t g(x)\,dx$$

for $t \ge a$.

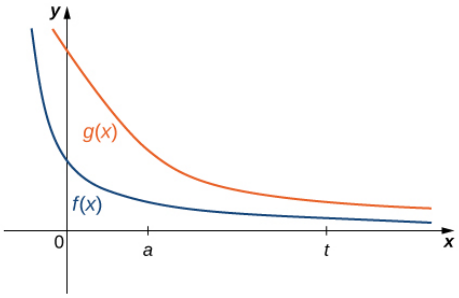


Figure 3.7.6: If $0 \le f(x) \le g(x)$ for $x \ge a$, then for $t \ge a$, $\int_a^t f(x)\,dx \le \int_a^t g(x)\,dx$.

Thus, if

$$\int_a^{+\infty} f(x)\,dx = \lim_{t\to+\infty} \int_a^t f(x)\,dx = +\infty,$$

then

$$\int_a^{+\infty} g(x)\,dx = \lim_{t\to+\infty} \int_a^t g(x)\,dx = +\infty$$

as well. That is, if the area of the region between the graph of $f(x)$ and the x-axis over $[a, +\infty)$ is infinite, then the area of the region between the graph of $g(x)$ and the x-axis over $[a, +\infty)$ is infinite too.

On the other hand, if

$$\int_a^{+\infty} g(x)\,dx = \lim_{t\to+\infty} \int_a^t g(x)\,dx = L$$

for some real number L, then

$$\int_a^{+\infty} f(x)\,dx = \lim_{t\to+\infty} \int_a^t f(x)\,dx$$

must converge to some value less than or equal to L, since $\int_a^t f(x)\,dx$ increases as t increases and $\int_a^t f(x)\,dx \le L$ for all $t \ge a$.

If the area of the region between the graph of $g(x)$ and the x-axis over $[a, +\infty)$ is finite, then the area of the region between the graph of $f(x)$ and the x-axis over $[a, +\infty)$ is also finite.

These conclusions are summarized in the following theorem.

A Comparison Theorem

Let $f(x)$ and $g(x)$ be continuous over $[a, +\infty)$. Assume that $0 \le f(x) \le g(x)$ for $x \ge a$.

i. If

$$\int_a^{+\infty} f(x)\,dx = \lim_{t\to+\infty} \int_a^t f(x)\,dx = +\infty,$$

then

$$\int_a^{+\infty} g(x)\,dx = \lim_{t\to+\infty} \int_a^t g(x)\,dx = +\infty.$$

ii. If

$$\int_a^{+\infty} g(x)\,dx = \lim_{t\to+\infty}\int_a^t g(x)\,dx = L,$$

where L is a real number, then

$$\int_a^{+\infty} f(x)\,dx = \lim_{t\to+\infty}\int_a^t f(x)\,dx = M$$

for some real number $M \le L$.

Example 3.7.9: Applying the Comparison Theorem

Use a comparison to show that $\displaystyle\int_1^{+\infty} \frac{1}{xe^x}\,dx$.

converges.

Solution

We can see that $0 \le \frac{1}{xe^x} \le \frac{1}{e^x} = e^{-x}$,

so if $\displaystyle\int_1^{+\infty} e^{-x}\,dx$ converges, then so does $\displaystyle\int_1^{+\infty} \frac{1}{xe^x}\,dx$.

To evaluate $\displaystyle\int_1^{+\infty} e^{-x}\,dx$, first rewrite it as a limit:

$$\begin{aligned}\int_1^{+\infty} e^{-x}\,dx &= \lim_{t\to+\infty}\int_1^t e^{-x}\,dx \\ &= \lim_{t\to+\infty}\left(-e^{-x}\right)\Big|_1^t \\ &= \lim_{t\to+\infty}\left(-e^{-t}+e^{-1}\right) \\ &= e^{-1}.\end{aligned}$$

Since $\displaystyle\int_1^{+\infty} e^{-x}\,dx$ converges, so does $\displaystyle\int_1^{+\infty} \frac{1}{xe^x}\,dx$.

Example 3.7.10: Applying the Comparison Theorem

Use the comparison theorem to show that $\displaystyle\int_1^{+\infty} \frac{1}{x^p}\,dx$ diverges for all $p < 1$.

Solution

For $p < 1$, $\dfrac{1}{x} \le \dfrac{1}{x^p}$ over $[1, +\infty)$.

In Example 3.7.1, we showed that $\displaystyle\int_1^{+\infty} \frac{1}{x}\,dx = +\infty$.

Therefore, $\displaystyle\int_1^{+\infty} \frac{1}{x^p}\,dx$ diverges for all $p < 1$.

Exercise 3.7.3

Use a comparison to show that $\displaystyle\int_e^{+\infty} \frac{\ln x}{x}\,dx$ diverges.

Hint

$\dfrac{1}{x} \le \dfrac{\ln x}{x}$ on $[e, +\infty)$

Answer

Since $\displaystyle\int_e^{+\infty} \frac{1}{x}\,dx = +\infty$, $\displaystyle\int_e^{+\infty} \frac{\ln x}{x}\,dx$ diverges.

Laplace Transforms

Laplace Transforms

In the last few chapters, we have looked at several ways to use integration for solving real-world problems. For this next project, we are going to explore a more advanced application of integration: integral transforms. Specifically, we describe the Laplace transform and some of its properties. The **Laplace transform** is used in engineering and physics to simplify the computations needed to solve some problems. It takes functions expressed in terms of time and *transforms* them to functions expressed in terms of frequency. It turns out that, in many cases, the computations needed to solve problems in the frequency domain are much simpler than those required in the time domain.

The Laplace transform is defined in terms of an integral as

$$\mathscr{L}\big(f(t)\big) = F(s) = \int_0^\infty e^{-st} f(t)\, dt.$$

Note that the input to a Laplace transform is a function of time, $f(t)$, and the output is a function of frequency, $F(s)$. Although many real-world examples require the use of complex numbers (involving the imaginary number $i = \sqrt{-1}$), in this project we limit ourselves to functions of real numbers.

Let's start with a simple example. Here we calculate the Laplace transform of $f(t) = t$. We have

$$\mathscr{L}(t) = \int_0^\infty te^{-st}\, dt.$$

This is an improper integral, so we express it in terms of a limit, which gives

$$\mathscr{L}(t) = \int_0^\infty te^{-st}\, dt = \lim_{z\to\infty} \int_0^z te^{-st}\, dt.$$

Now we use integration by parts to evaluate the integral. Note that we are integrating with respect to t, so we treat the variable s as a constant. We have

$$\begin{aligned} u &= t & dv &= e^{-st}\, dt \\ du &= dt & v &= -\tfrac{1}{s}e^{-st} \end{aligned}$$

Then we obtain

$$\begin{aligned} \lim_{z\to\infty} \int_0^z te^{-st}\, dt &= \lim_{z\to\infty} \left[-\frac{t}{s}e^{-st}\Big|_0^z + \frac{1}{s}\int_0^z e^{-st}\, dt \right] \\ &= \lim_{z\to\infty} \left[\left[-\frac{z}{s}e^{-sz} + \frac{0}{s}e^{-0s} \right] + \frac{1}{s}\int_0^z e^{-st}\, dt \right] \\ &= \lim_{z\to\infty} \left[\left[-\frac{z}{s}e^{-sz} + 0 \right] - \frac{1}{s}\left[\frac{e^{-st}}{s} \right] \Big|_0^z \right] \\ &= \lim_{z\to\infty} \left[-\frac{z}{s}e^{-sz} - \frac{1}{s^2}\left[e^{-sz} - 1 \right] \right] \\ &= \lim_{z\to\infty} \left[-\frac{z}{se^{sz}} \right] - \lim_{z\to\infty} \frac{1}{s^2 e^{sz}} + \lim_{z\to\infty} \frac{1}{s^2} \\ &= 0 - 0 + \frac{1}{s^2} \qquad \text{Note that the first limit requires L'Hôpital's rule. See work below.} \\ &= \frac{1}{s^2}. \end{aligned}$$

As noted above, $\lim\limits_{z\to\infty} \left[-\dfrac{z}{se^{sz}} \right]$ is indeterminate of the form $-\dfrac{\infty}{\infty}$ and requires L'Hôpital's rule to evaluate. Using L'Hôpital's rule,

$$\lim_{z\to\infty} \left[-\frac{z}{se^{sz}} \right] = \lim_{z\to\infty} \left[-\frac{1}{s^2 e^{sz}} \right] = 0.$$

1. Calculate the Laplace transform of $f(t) = 1$.
2. Calculate the Laplace transform of $f(t) = e^{-3t}$.
3. Calculate the Laplace transform of $f(t) = t^2$. (Note, you will have to integrate by parts twice.)

Laplace transforms are often used to solve differential equations. Differential equations are not covered in detail until later in this book; but, for now, let's look at the relationship between the Laplace transform of a function and the Laplace transform of its derivative.

Let's start with the definition of the Laplace transform. We have

$$\mathscr{L}\big(f(t)\big) = \int_0^{\infty} e^{-st} f(t)\, dt = \lim_{z\to\infty} \int_0^{z} e^{-st} f(t)\, dt.$$

Use integration by parts to evaluate $\lim_{z\to\infty} \int_0^{z} e^{-st} f(t)\, dt$. (Let $u = f(t)$ and $dv = e^{-st}\, dt$.)

After integrating by parts and evaluating the limit, you should see that

$$\mathscr{L}\big(f(t)\big) = \frac{f(0)}{s} + \frac{1}{s}\big[\mathscr{L}\big(f'(t)\big)\big].$$

Then,

$$\mathscr{L}\big(f'(t)\big) = s\mathscr{L}\big(f(t)\big) - f(0).$$

Thus, differentiation in the time domain simplifies to multiplication by s in the frequency domain.

The final thing we look at in this project is how the Laplace transforms of $f(t)$ and its antiderivative are related. Let $g(t) = \int_0^{t} f(u)\, du$.

Then,

$$\mathscr{L}\big(g(t)\big) = \int_0^{\infty} e^{-st} g(t)\, dt = \lim_{z\to\infty} \int_0^{z} e^{-st} g(t)\, dt.$$

Use integration by parts to evaluate $\lim_{z\to\infty} \int_0^{z} e^{-st} g(t)\, dt$. (Let $u = g(t)$ and $dv = e^{-st}\, dt$. Note, by the way that we have defined $g(t)$, $du = f(t)\, dt$.)

As you might expect, you should see that

$$\mathscr{L}\big(g(t)\big) = \frac{1}{s} \cdot \mathscr{L}\big(f(t)\big).$$

Integration in the time domain simplifies to division by s in the frequency domain.

Key Concepts

- Integrals of functions over infinite intervals are defined in terms of limits.
- Integrals of functions over an interval for which the function has a discontinuity at an endpoint may be defined in terms of limits.
- The convergence or divergence of an improper integral may be determined by comparing it with the value of an improper integral for which the convergence or divergence is known.

Key Equations

- **Improper integrals**

$$\int_a^{+\infty} f(x)\, dx = \lim_{t\to+\infty} \int_a^{t} f(x)\, dx$$

$$\int_{-\infty}^{b} f(x)\, dx = \lim_{t\to-\infty} \int_t^{b} f(x)\, dx$$

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_0^{+\infty} f(x)\, dx$$

Glossary

improper integral

an integral over an infinite interval or an integral of a function containing an infinite discontinuity on the interval; an improper integral is defined in terms of a limit. The improper integral converges if this limit is a finite real number; otherwise, the improper integral diverges

This page titled 3.7: Improper Integrals is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **7.7: Improper Integrals** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

CHAPTER OVERVIEW

4: Introduction to Differential Equations

Many real-world phenomena can be modeled mathematically by using differential equations. Population growth, radioactive decay, predator-prey models, and spring-mass systems are four examples of such phenomena. In this chapter we study some of these applications. A goal of this chapter is to develop solution techniques for different types of differential equations. As the equations become more complicated, the solution techniques also become more complicated, and in fact an entire course could be dedicated to the study of these equations. In this chapter we study several types of differential equations and their corresponding methods of solution.

4.E: Introduction to Differential Equations (Exercises)
4.0: Prelude to Differential Equations
4.1: Basics of Differential Equations
4.2: Direction Fields and Numerical Methods
4.3: Separable Equations
4.4: The Logistic Equation

Thumbnail: An exponential growth model of population.

Contributors

Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 4: Introduction to Differential Equations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.E: Introduction to Differential Equations (Exercises)

This page titled 4.E: Introduction to Differential Equations (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.0: Prelude to Differential Equations

Many real-world phenomena can be modeled mathematically by using differential equations. Population growth, radioactive decay, predator-prey models, and spring-mass systems are four examples of such phenomena. In this chapter we study some of these applications. Suppose we wish to study a population of deer over time and determine the total number of animals in a given area. We can first observe the population over a period of time, estimate the total number of deer, and then use various assumptions to derive a mathematical model for different scenarios. Some factors that are often considered are environmental impact, threshold population values, and predators. In this chapter we see how differential equations can be used to predict populations over time.



Figure 4.0.1: The white-tailed deer (Odocoileus virginianus) of the eastern United States. Differential equations can be used to study animal populations. (credit: modification of work by Rachel Kramer, Flickr)

Another goal of this chapter is to develop solution techniques for different types of differential equations. As the equations become more complicated, the solution techniques also become more complicated, and in fact an entire course could be dedicated to the study of these equations. In this chapter we study several types of differential equations and their corresponding methods of solution.

This page titled 4.0: Prelude to Differential Equations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **8.0: Prelude to Differential Equations** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

4.1: Basics of Differential Equations

Learning Objectives

- Identify the order of a differential equation.
- Explain what is meant by a solution to a differential equation.
- Distinguish between the general solution and a particular solution of a differential equation.
- Identify an initial-value problem.
- Identify whether a given function is a solution to a differential equation or an initial-value problem.

Calculus is the mathematics of change, and rates of change are expressed by derivatives. Thus, one of the most common ways to use calculus is to set up an equation containing an unknown function $y = f(x)$ and its derivative, known as a differential equation. Solving such equations often provides information about how quantities change and frequently provides insight into how and why the changes occur.

Techniques for solving differential equations can take many different forms, including direct solution, use of graphs, or computer calculations. We introduce the main ideas in this chapter and describe them in a little more detail later in the course. In this section we study what differential equations are, how to verify their solutions, some methods that are used for solving them, and some examples of common and useful equations.

General Differential Equations

Consider the equation $y' = 3x^2$, which is an example of a differential equation because it includes a derivative. There is a relationship between the variables x and y : y is an unknown function of x. Furthermore, the left-hand side of the equation is the derivative of y. Therefore we can interpret this equation as follows: Start with some function $y = f(x)$ and take its derivative. The answer must be equal to $3x^2$. What function has a derivative that is equal to $3x^2$? One such function is $y = x^3$, so this function is considered a *solution* to a differential equation.

Definition: differential equation

A *differential equation* is an equation involving an unknown function $y = f(x)$ and one or more of its derivatives. A solution to a differential equation is a function $y = f(x)$ that satisfies the differential equation when f and its derivatives are substituted into the equation.

Go to this website to explore more on this topic.

Some examples of differential equations and their solutions appear in Table 4.1.1.

Table 4.1.1: Examples of Differential Equations and Their Solutions

Equation	Solution
$y' = 2x$	$y = x^2$
$y' + 3y = 6x + 11$	$y = e^{-3x} + 2x + 3$
$y'' - 3y' + 2y = 24e^{-2x}$	$y = 3e^x - 4e^{2x} + 2e^{-2x}$

Note that a solution to a differential equation is not necessarily unique, primarily because the derivative of a constant is zero. For example, $y = x^2 + 4$ is also a solution to the first differential equation in Table. We will return to this idea a little bit later in this section. For now, let's focus on what it means for a function to be a solution to a differential equation.

Example 4.1.1: Verifying Solutions of Differential Equations

Verify that the function $y = e^{-3x} + 2x + 3$ is a solution to the differential equation $y' + 3y = 6x + 11$.

Solution

To verify the solution, we first calculate y' using the chain rule for derivatives. This gives $y' = -3e^{-3x} + 2$. Next we substitute y and y' into the left-hand side of the differential equation:

$$(-3e^{-2x} + 2) + 3(e^{-2x} + 2x + 3).$$

The resulting expression can be simplified by first distributing to eliminate the parentheses, giving

$$-3e^{-2x}+2+3e^{-2x}+6x+9.$$

Combining like terms leads to the expression $6x+11$, which is equal to the right-hand side of the differential equation. This result verifies that $y=e^{-3x}+2x+3$ is a solution of the differential equation.

Exercise 4.1.1

Verify that $y=2e^{3x}-2x-2$ is a solution to the differential equation $y'-3y=6x+4$.

Hint

First calculate y' then substitute both y' and y into the left-hand side.

It is convenient to define characteristics of differential equations that make it easier to talk about them and categorize them. The most basic characteristic of a differential equation is its order.

Definition: order of a differential equation

The *order of a differential equation* is the highest order of any derivative of the unknown function that appears in the equation.

Example 4.1.2: Identifying the Order of a Differential Equation

The highest derivative in the equation is y',

What is the order of each of the following differential equations?

a. $y'-4y=x^2-3x+4$
b. $x^2y'''-3xy''+xy'-3y=\sin x$
c. $\frac{4}{x}y^{(4)}-\frac{6}{x^2}y''+\frac{12}{x^4}y=x^3-3x^2+4x-12$

Solution:

a. The highest derivative in the equation is y',so the order is 1.
b. The highest derivative in the equation is y''', so the order is 3.
c. The highest derivative in the equation is $y^{(4)}$, so the order is 4.

Exercise 4.1.2

What is the order of the following differential equation?

$$(x^4-3x)y^{(5)}-(3x^2+1)y'+3y=\sin x\cos x$$

Hint

What is the highest derivative in the equation?

Answer

5

General and Particular Solutions

We already noted that the differential equation $y'=2x$ has at least two solutions: $y=x^2$ and $y=x^2+4$. The only difference between these two solutions is the last term, which is a constant. What if the last term is a different constant? Will this expression still be a solution to the differential equation? In fact, any function of the form $y=x^2+C$, where C represents any constant, is a solution as well. The reason is that the derivative of x^2+C is $2x$, regardless of the value of C. It can be shown that any solution of this differential equation must be of the form $y=x^2+C$. This is an example of a **general solution** to a differential equation. A graph of some of these solutions is given in Figure. (Note: in this graph we used even integer values for C ranging between -4 and 4. In fact, there is no restriction on the value of C; it can be an integer or not.)

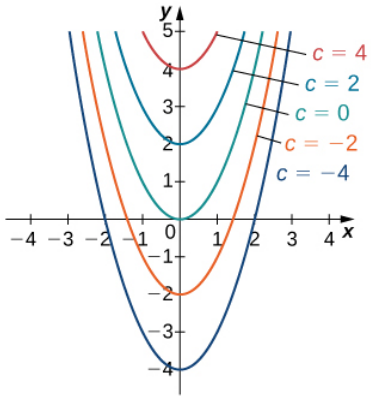


Figure 4.1.1: Family of solutions to the differential equation $y' = 2x$.

In this example, we are free to choose any solution we wish; for example, $y = x^2 - 3$ is a member of the family of solutions to this differential equation. This is called a particular solution to the differential equation. A particular solution can often be uniquely identified if we are given additional information about the problem.

Example 4.1.3: Finding a Particular Solution

Find the particular solution to the differential equation $y' = 2x$ passing through the point $(2, 7)$.

Solution

Any function of the form $y = x^2 + C$ is a solution to this differential equation. To determine the value of C, we substitute the values $x = 2$ and $y = 7$ into this equation and solve for C:

$$\begin{aligned} y &= x^2 + C \\ 7 &= 2^2 + C \\ &= 4 + C \\ C &= 3. \end{aligned}$$

Therefore the particular solution passing through the point $(2, 7)$ is $y = x^2 + 3$.

Exercise 4.1.3

Find the particular solution to the differential equation

$$y' = 4x + 3$$

passing through the point $(1, 7)$, given that $y = 2x^2 + 3x + C$ is a general solution to the differential equation.

Hint

First substitute $x = 1$ and $y = 7$ into the equation, then solve for C.

Answer

$$y = 2x^2 + 3x + 2$$

Initial-Value Problems

Usually a given differential equation has an infinite number of solutions, so it is natural to ask which one we want to use. To choose one solution, more information is needed. Some specific information that can be useful is an *initial value*, which is an ordered pair that is used to find a particular solution.

A differential equation together with one or more initial values is called an *initial-value problem*. The general rule is that the number of initial values needed for an initial-value problem is equal to the order of the differential equation. For example, if we have the differential equation $y' = 2x$, then $y^{(3)} = 7$ is an initial value, and when taken together, these equations form an initial-value problem. The differential equation $y'' - 3y' + 2y = 4e^x$ is second order, so we need two initial values. With initial-value

problems of order greater than one, the same value should be used for the independent variable. An example of initial values for this second-order equation would be $y(0)=2$ and $y'(0)=-1$. These two initial values together with the differential equation form an initial-value problem. These problems are so named because often the independent variable in the unknown function is t, which represents time. Thus, a value of $t=0$ represents the beginning of the problem.

Example 4.1.4: Verifying a Solution to an Initial-Value Problem

Verify that the function $y=2e^{-2t}+e^{t}$ is a solution to the initial-value problem

$$y'+2y=3e^{t}, \quad y(0)=3.$$

Solution

For a function to satisfy an initial-value problem, it must satisfy both the differential equation and the initial condition. To show that y satisfies the differential equation, we start by calculating y'. This gives $y'=-4e^{-2t}+e^{t}$. Next we substitute both y and y' into the left-hand side of the differential equation and simplify:

$$\begin{aligned} y'+2y &= (-4e^{-2t}+e^{t})+2(2e^{-2t}+e^{t}) \\ &= -4e^{-2t}+e^{t}+4e^{-2t}+2e^{t}=3e^{t}. \end{aligned}$$

This is equal to the right-hand side of the differential equation, so $y=2e^{-2t}+e^{t}$ solves the differential equation. Next we calculate $y(0)$:

$$y(0)=2e^{-2(0)}+e^{0}=2+1=3.$$

This result verifies the initial value. Therefore the given function satisfies the initial-value problem.

Exercise 4.1.4

Verify that $y=3e^{2t}+4\sin t$ is a solution to the initial-value problem

$$y'-2y=4\cos t-8\sin t, y(0)=3.$$

Hint

First verify that y solves the differential equation. Then check the initial value.

In Example, the initial-value problem consisted of two parts. The first part was the differential equation $y'+2y=3e^{x}$, and the second part was the initial value $y(0)=3$. These two equations together formed the initial-value problem.

The same is true in general. An initial-value problem will consists of two parts: the differential equation and the initial condition. The differential equation has a family of solutions, and the initial condition determines the value of C. The family of solutions to the differential equation in Example is given by $y=2e^{-2t}+Ce^{t}$. This family of solutions is shown in Figure, with the particular solution $y=2e^{-2t}+e^{t}$ labeled.

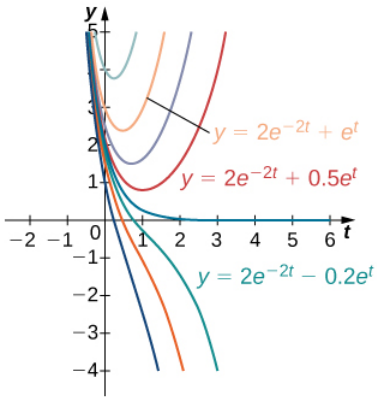


Figure 4.1.2: A family of solutions to the differential equation $y'+2y=3e^{t}$. The particular solution $y=2e^{-2t}+e^{t}$ is labeled.

Example 4.1.5: Solving an Initial-value Problem

Solve the following initial-value problem:

$$y' = 3e^x + x^2 - 4, y(0) = 5.$$

Solution

The first step in solving this initial-value problem is to find a general family of solutions. To do this, we find an antiderivative of both sides of the differential equation

$$\int y' dx = \int (3e^x + x^2 - 4)\, dx,$$

namely,

$$y + C_1 = 3e^x + \tfrac{1}{3}x^3 - 4x + C_2 \quad .$$

We are able to integrate both sides because the y term appears by itself. Notice that there are two integration constants: C_1 and C_2. Solving Equation for y gives

$$y = 3e^x + \tfrac{1}{3}x^3 - 4x + C_2 - C_1.$$

Because C_1 and C_2 are both constants, $C_2 - C_1$ is also a constant. We can therefore define $C = C_2 - C_1$, which leads to the equation

$$y = 3e^x + \tfrac{1}{3}x^3 - 4x + C.$$

Next we determine the value of C. To do this, we substitute $x = 0$ and $y = 5$ into Equation and solve for C:

$$5 = 3e^0 + \tfrac{1}{3}0^3 - 4(0) + C$$

$$5 = 3 + C$$

$$C = 2.$$

Now we substitute the value $C = 2$ into Equation. The solution to the initial-value problem is $y = 3e^x + \frac{1}{3}x^3 - 4x + 2$.

Analysis

The difference between a general solution and a particular solution is that a general solution involves a family of functions, either explicitly or implicitly defined, of the independent variable. The initial value or values determine which particular solution in the family of solutions satisfies the desired conditions.

Exercise 4.1.5

Solve the initial-value problem

$$y' = x^2 - 4x + 3 - 6e^x, y(0) = 8.$$

Hint

First take the antiderivative of both sides of the differential equation. Then substitute $x = 0$ and $y = 8$ into the resulting equation and solve for C.

Answer

$y = \frac{1}{3}x^3 - 2x^2 + 3x - 6e^x + 14$

In physics and engineering applications, we often consider the forces acting upon an object, and use this information to understand the resulting motion that may occur. For example, if we start with an object at Earth's surface, the primary force acting upon that object is gravity. Physicists and engineers can use this information, along with Newton's **second law of motion** (in equation form $F = ma$, where F represents force, m represents mass, and a represents acceleration), to derive an equation that can be solved.

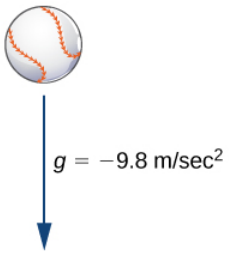


Figure 4.1.3: For a baseball falling in air, the only force acting on it is gravity (neglecting air resistance).

In Figure 4.1.3 we assume that the only force acting on a baseball is the force of gravity. This assumption ignores air resistance. (The force due to air resistance is considered in a later discussion.) The acceleration due to gravity at Earth's surface, g, is approximately $9.8\ \text{m/s}^2$. We introduce a frame of reference, where Earth's surface is at a height of 0 meters. Let $v(t)$ represent the velocity of the object in meters per second. If $v(t) > 0$, the ball is rising, and if $v(t) < 0$, the ball is falling (Figure).



Figure 4.1.4: Possible velocities for the rising/falling baseball.

Our goal is to solve for the velocity $v(t)$ at any time t. To do this, we set up an initial-value problem. Suppose the mass of the ball is m, where m is measured in kilograms. We use Newton's second law, which states that the force acting on an object is equal to its mass times its acceleration $(F = ma)$. Acceleration is the derivative of velocity, so $a(t) = v'(t)$. Therefore the force acting on the baseball is given by $F = mv'(t)$. However, this force must be equal to the force of gravity acting on the object, which (again using Newton's second law) is given by $F_g = -mg$, since this force acts in a downward direction. Therefore we obtain the equation $F = F_g$, which becomes $mv'(t) = -mg$. Dividing both sides of the equation by m gives the equation

$$v'(t) = -g. \tag{4.1.1}$$

Notice that this differential equation remains the same regardless of the mass of the object.

We now need an initial value. Because we are solving for velocity, it makes sense in the context of the problem to assume that we know the **initial velocity**, or the velocity at time $t = 0$. This is denoted by $v(0) = v_0$.

Example 4.1.6: Velocity of a Moving Baseball

A baseball is thrown upward from a height of 3 meters above Earth's surface with an initial velocity of 10 m/s, and the only force acting on it is gravity. The ball has a mass of 0.15 kg at Earth's surface.

a. Find the velocity $v(t)$ of the basevall at time t.
b. What is its velocity after 2 seconds?

Solution

a. From the preceding discussion, the differential equation that applies in this situation is

$$v'(t) = -g,$$

where $g = 9.8\ \text{m/s}^2$. The initial condition is $v(0) = v_0$, where $v_0 = 10$ m/s. Therefore the initial-value problem is $v'(t) = -9.8\ \text{m/s}^2$, $v(0) = 10$ m/s.

The first step in solving this initial-value problem is to take the antiderivative of both sides of the differential equation. This gives

$$\int v'(t)\,dt = \int -9.8\,dt$$

$$v(t) = -9.8t + C.$$

The next step is to solve for C. To do this, substitute $t = 0$ and $v(0) = 10$:

$$v(t) = -9.8t + C$$

$$v(0) = -9.8(0) + C$$

$$10 = C.$$

Therefore $C = 10$ and the velocity function is given by $v(t) = -9.8t + 10$.

b. To find the velocity after 2 seconds, substitute $t = 2$ into $v(t)$.

$$v(t) = -9.8t + 10$$

$$v(2) = -9.8(2) + 10$$

$$v(2) = -9.6.$$

The units of velocity are meters per second. Since the answer is negative, the object is falling at a speed of 9.6 m/s.

Exercise 4.1.6

Suppose a rock falls from rest from a height of 100 meters and the only force acting on it is gravity. Find an equation for the velocity $v(t)$ as a function of time, measured in meters per second.

Hint

What is the initial velocity of the rock? Use this with the differential equation in Example $4.1.6$ to form an initial-value problem, then solve for $v(t)$.

Answer

$v(t) = -9.8t$

A natural question to ask after solving this type of problem is how high the object will be above Earth's surface at a given point in time. Let $s(t)$ denote the height above Earth's surface of the object, measured in meters. Because velocity is the derivative of position (in this case height), this assumption gives the equation $s'(t) = v(t)$. An initial value is necessary; in this case the initial height of the object works well. Let the initial height be given by the equation $s(0) = s_0$. Together these assumptions give the initial-value problem

$$s'(t) = v(t), s(0) = s_0. \tag{4.1.2}$$

If the velocity function is known, then it is possible to solve for the position function as well.

Example 4.1.7: Height of a Moving Baseball

A baseball is thrown upward from a height of 3 meters above Earth's surface with an initial velocity of $10m/s$, and the only force acting on it is gravity. The ball has a mass of 0.15 kilogram at Earth's surface.

a. Find the position $s(t)$ of the baseball at time t.
b. What is its height after 2 seconds?

Solution

We already know the velocity function for this problem is $v(t) = -9.8t + 10$. The initial height of the baseball is 3 meters, so $s_0 = 3$. Therefore the initial-value problem for this example is

To solve the initial-value problem, we first find the antiderivatives:

$$\int s'(t)\,dt = \int (-9.8t + 10)\,dt$$

$$s(t) = -4.9t^2 + 10t + C.$$

Next we substitute $t = 0$ and solve for C:

$$s(t) = -4.9t^2 + 10t + C$$

$$s(0) = -4.9(0)^2 + 10(0) + C$$

$$3 = C.$$

Therefore the position function is $s(t) = -4.9t^2 + 10t + 3$.

b. The height of the baseball after 2 sec is given by $s(2)$:

$$s(2) = -4.9(2)^2 + 10(2) + 3 = -4.9(4) + 23 = 3.4.$$

Therefore the baseball is 3.4 meters above Earth's surface after 2 seconds. It is worth noting that the mass of the ball cancelled out completely in the process of solving the problem.

Key Concepts

- A differential equation is an equation involving a function $y = f(x)$ and one or more of its derivatives. A solution is a function $y = f(x)$ that satisfies the differential equation when f and its derivatives are substituted into the equation.
- The order of a differential equation is the highest order of any derivative of the unknown function that appears in the equation.
- A differential equation coupled with an initial value is called an initial-value problem. To solve an initial-value problem, first find the general solution to the differential equation, then determine the value of the constant. Initial-value problems have many applications in science and engineering.

Glossary

differential equation
an equation involving a function $y = y(x)$ and one or more of its derivatives

general solution (or family of solutions)
the entire set of solutions to a given differential equation

initial value(s)
a value or set of values that a solution of a differential equation satisfies for a fixed value of the independent variable

initial velocity
the velocity at time $t = 0$

initial-value problem
a differential equation together with an initial value or values

order of a differential equation
the highest order of any derivative of the unknown function that appears in the equation

particular solution
member of a family of solutions to a differential equation that satisfies a particular initial condition

solution to a differential equation
a function $y = f(x)$ that satisfies a given differential equation

Contributors

- Gilbert Strang (MIT) and Edwin "Jed" Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 4.1: Basics of Differential Equations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.2: Direction Fields and Numerical Methods

Learning Objectives

- Draw the direction field for a given first-order differential equation.
- Use a direction field to draw a solution curve of a first-order differential equation.
- Use Euler's Method to approximate the solution to a first-order differential equation.

For the rest of this chapter we will focus on various methods for solving differential equations and analyzing the behavior of the solutions. In some cases it is possible to predict properties of a solution to a differential equation without knowing the actual solution. We will also study numerical methods for solving differential equations, which can be programmed by using various computer languages or even by using a spreadsheet program, such as Microsoft Excel.

Creating Direction Fields

Direction fields (also called slope fields) are useful for investigating first-order differential equations. In particular, we consider a first-order differential equation of the form

$$y' = f(x, y).$$

An applied example of this type of differential equation appears in Newton's law of cooling, which we will solve explicitly later in this chapter. First, though, let us create a direction field for the differential equation

$$T'(t) = -0.4(T - 72).$$

Here $T(t)$ represents the temperature (in degrees Fahrenheit) of an object at time t, and the ambient temperature is $72^\circ F$. Figure $4.2.1$ shows the direction field for this equation.

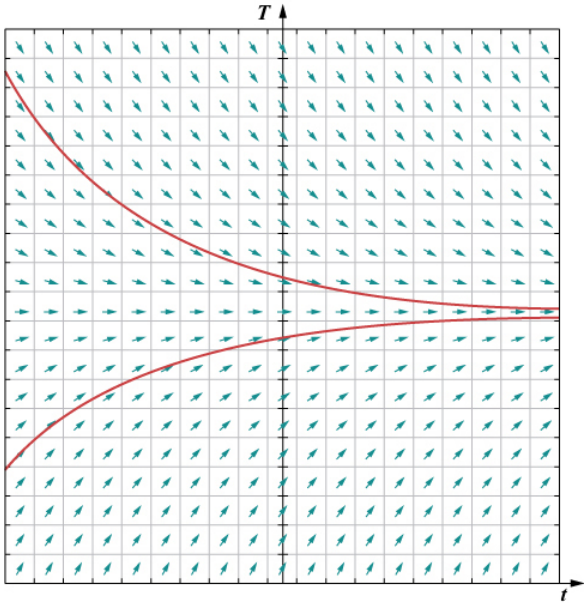


Figure 4.2.1: Direction field for the differential equation $T'(t) = -0.4(T - 72)$. Two solutions are plotted: one with initial temperature less than $72^\circ F$ and the other with initial temperature greater than $72^\circ F$.

The idea behind a direction field is the fact that the derivative of a function evaluated at a given point is the slope of the tangent line to the graph of that function at the same point. Other examples of differential equations for which we can create a direction field include

$$y' = 3x + 2y - 4$$

$$y' = x^2 - y^2$$

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

To create a direction field, we start with the first equation: $y' = 3x + 2y - 4$. We let (x_0, y_0) be any ordered pair, and we substitute these numbers into the right-hand side of the differential equation. For example, if we choose $x = 1$ and $y = 2$, substituting into the right-hand side of the differential equation yields

$$y' = 3x + 2y - 4 = 3(1) + 2(2) - 4 = 3.$$

This tells us that if a solution to the differential equation $y' = 3x + 2y - 4$ passes through the point $(1, 2)$, then the slope of the solution at that point must equal 3. To start creating the direction field, we put a short line segment at the point $(1, 2)$ having slope 3. We can do this for any point in the domain of the function $f(x, y) = 3x + 2y - 4$, which consists of all ordered pairs (x, y) in R^2. Therefore any point in the Cartesian plane has a slope associated with it, assuming that a solution to the differential equation passes through that point. The direction field for the differential equation $y' = 3x + 2y - 4$ is shown in Figure 4.2.2.

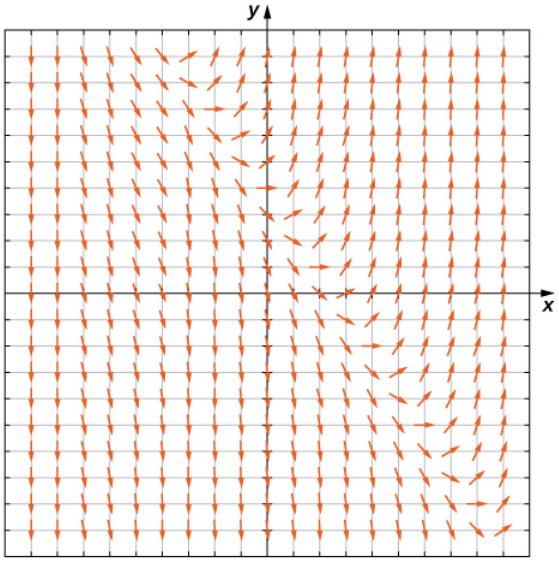


Figure 4.2.2: Direction field for the differential equation $y' = 3x + 2y - 4$.

We can generate a direction field of this type for any differential equation of the form $y' = f(x, y)$.

Definition: Direction Field (Slope Field)

A direction field (slope field) is a mathematical object used to graphically represent solutions to a first-order differential equation. At each point in a direction field, a line segment appears whose slope is equal to the slope of a solution to the differential equation passing through that point.

Using Direction Fields

We can use a direction field to predict the behavior of solutions to a differential equation without knowing the actual solution. For example, the direction field in Figure 4.2.3 serves as a guide to the behavior of solutions to the differential equation $y' = 3x + 2y - 4$.

To use a direction field, we start by choosing any point in the field. The line segment at that point serves as a signpost telling us what direction to go from there. For example, if a solution to the differential equation passes through the point $(0, 1)$, then the slope of the solution passing through that point is given by $y' = 3(0) + 2(1) - 4 = -2$. Now let x increase slightly, say to $x = 0.1$. Using the method of linear approximations gives a formula for the approximate value of y for $x = 0.1$. In particular,

$$L(x) = y_0 + f'(x_0)(x - x_0) = 1 - 2(x_0 - 0) = 1 - 2x_0.$$

Substituting $x_0 = 0.1$ into $L(x)$ gives an approximate y value of 0.8.

At this point the slope of the solution changes (again according to the differential equation). We can keep progressing, recalculating the slope of the solution as we take small steps to the right, and watching the behavior of the solution. Figure 4.2.3 shows a graph

of the solution passing through the point $(0, 1)$.

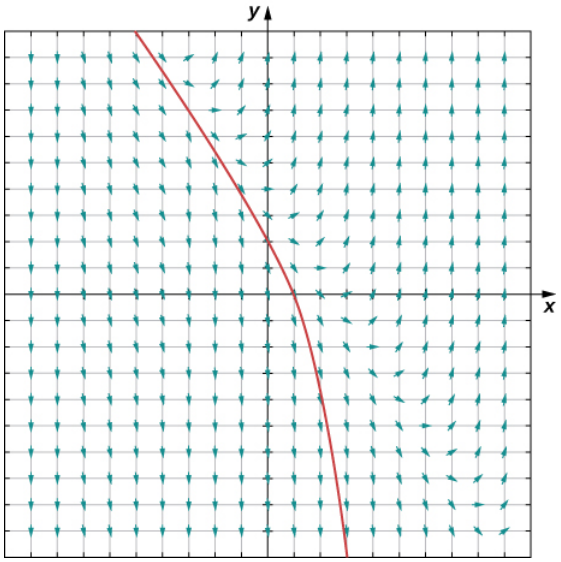


Figure 4.2.3: Direction field for the differential equation $y' = 3x + 2y - 4$ with the solution passing through the point $(0, 1)$.

The curve is the graph of the solution to the initial-value problem

$$y' = 3x + 2y - 4,\ y(0) = 1.$$

This curve is called a solution curve passing through the point $(0, 1)$. The exact solution to this initial-value problem is

$$y = -\frac{3}{2}x + \frac{5}{4} - \frac{1}{4}e^{2x},$$

and the graph of this solution is identical to the curve in Figure 4.2.3.

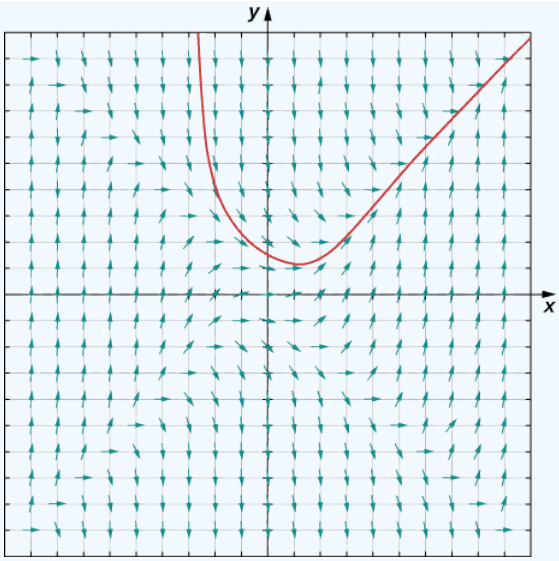
? Exercise 4.2.1

Create a direction field for the differential equation $y' = x^2 - y^2$ and sketch a solution curve passing through the point $(-1, 2)$.

Hint

Use x and y values ranging from -5 to 5. For each coordinate pair, calculate y' using the right-hand side of the differential equation.

Answer



Now consider the direction field for the differential equation $y' = (x-3)(y^2-4)$, shown in Figure 4.2.4. This direction field has several interesting properties. First of all, at $y = -2$ and $y = 2$, horizontal dashes appear all the way across the graph. This means that if $y = -2$, then $y' = 0$. Substituting this expression into the right-hand side of the differential equation gives

$$\begin{aligned}(x-3)(y^2-4) &= (x-3)((-2)^2-4)\\ &= (x-3)(0)\\ &= 0\\ &= y'.\end{aligned}$$

Therefore $y = -2$ is a solution to the differential equation. Similarly, $y = 2$ is a solution to the differential equation. These are the only constant-valued solutions to the differential equation, as we can see from the following argument. Suppose $y = k$ is a constant solution to the differential equation. Then $y' = 0$. Substituting this expression into the differential equation yields $0 = (x-3)(k^2-4)$. This equation must be true for all values of x, so the second factor must equal zero. This result yields the equation $k^2 - 4 = 0$. The solutions to this equation are $k = -2$ and $k = 2$, which are the constant solutions already mentioned. These are called the equilibrium solutions to the differential equation.

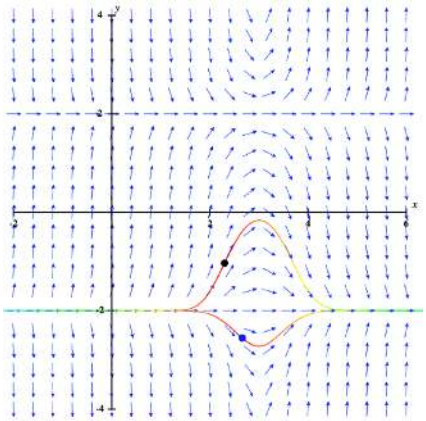


Figure 4.2.4: Direction field for the differential equation $y' = (x-3)(y^2-4)$ showing two solutions. These solutions are very close together, but one is barely above the equilibrium solution $y = -2$ and the other is barely below the same equilibrium solution.

Definition: Equilibrium Solutions

Consider the differential equation $y' = f(x, y)$. An *equilibrium solution* is any solution to the differential equation of the form $y = c$, where c is a constant.

To determine the equilibrium solutions to the differential equation $y' = f(x, y)$, set the right-hand side equal to zero. An equilibrium solution of the differential equation is any function of the form $y = k$ such that $f(x, k) = 0$ for all values of x in the domain of f.

An important characteristic of equilibrium solutions concerns whether or not they approach the line $y = k$ as an asymptote for large values of x.

Definition: asymptotically Stable, Unstable and Semi-Stable Solutions

Consider the differential equation $y' = f(x, y)$, and assume that all solutions to this differential equation are defined for $x \geq x_0$. Let $y = k$ be an equilibrium solution to the differential equation.

1. $y = k$ is an *asymptotically stable solution* to the differential equation if there exists $\varepsilon > 0$ such that for any value $c \in (k - \varepsilon,\ k + \varepsilon)$ the solution to the initial-value problem $y' = f(x, y),\ y(x_0) = c$ approaches k as x approaches infinity.
2. $y = k$ is an *asymptotically unstable solution* to the differential equation if there exists $\varepsilon > 0$ such that for any value $c \in (k - \varepsilon,\ k + \varepsilon)$ the solution to the initial-value problem $y' = f(x, y),\ y(x_0) = c$ never approaches k as x approaches infinity.
3. $y = k$ is an *asymptotically semi-stable solution* to the differential equation if it is neither asymptotically stable nor asymptotically unstable.

Now we return to the differential equation $y' = (x - 3)(y^2 - 4)$, with the initial condition $y(0) = 0.5$. The direction field for this initial-value problem, along with the corresponding solution, is shown in Figure $4.2.5$.

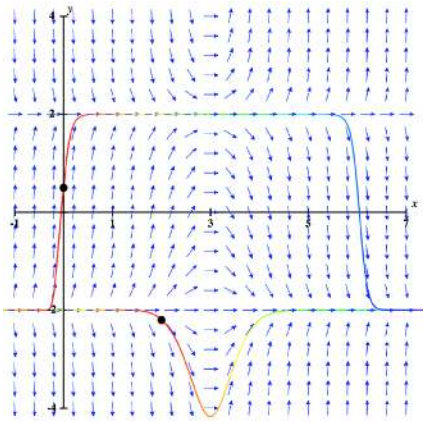


Figure 4.2.5: Direction field for the initial-value problem $y' = (x - 3)(y^2 - 4),\ y(0) = 0.5$. A second solution curve is graphed with initial condition $(2, -2.2)$.

The values of the solution to this initial-value problem stay between $y = -2$ and $y = 2$, which are the equilibrium solutions to the differential equation. Furthermore, as x approaches infinity, although they initially appear to approach the line y = 2, the y-coordinates clearly approach -2. The behavior of solutions is similar if the initial value is below -2, for example, $y(2) = -2.2$. In this case, the solutions increase and approach $y = -2$ as x approaches infinity. Therefore $y = -2$ is an asymptotically stable solution to this differential equation.

What happens when the initial value is above $y = 2$? This scenario is illustrated in Figure $4.2.6$, with the initial value $y(0) = 3$.

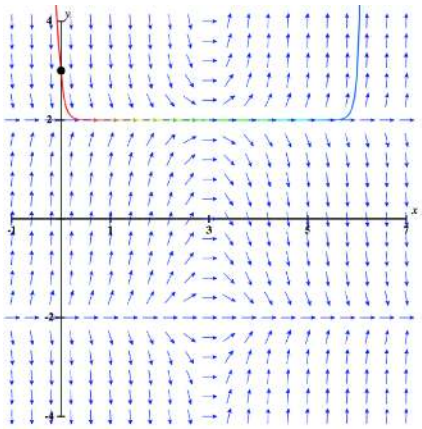


Figure 4.2.6: Direction field for the initial-value problem $y' = (x-3)(y^2-4),\ y(0) = 3$.

The solution increases rapidly toward positive infinity as x approaches infinity. Furthermore, if the initial value is slightly below 2, then the solution approaches -2, which is the other equilibrium solution. Therefore in neither case does the solution approach $y = 2$, so $y = 2$ is called an asymptotically unstable, or unstable, equilibrium solution.

Example 4.2.1: Stability of an Equilibrium Solution

Create a direction field for the differential equation $y' = (y-3)^2(y^2+y-2)$ and identify any equilibrium solutions. Classify each of the equilibrium solutions as stable, unstable, or semi-stable.

Solution

The direction field is shown in Figure 4.2.7.

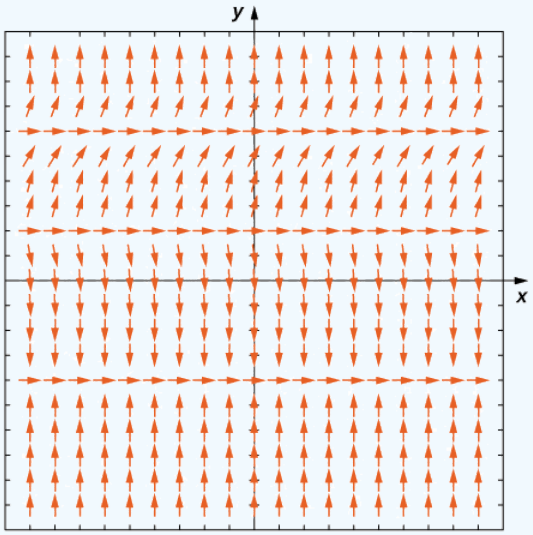


Figure 4.2.7: Direction field for the differential equation $y' = (y-3)^2(y^2+y-2)$.

The equilibrium solutions are $y = -2,\ y = 1,$ and $y = 3$. To classify each of the solutions, look at an arrow directly above or below each of these values. For example, at $y = -2$ the arrows directly below this solution point up, and the arrows directly above the solution point down. Therefore all initial conditions close to $y = -2$ approach $y = -2$, and the solution is stable. For the solution $y = 1$, all initial conditions above and below $y = 1$ are repelled (pushed away) from $y = 1$, so this solution is unstable. The solution $y = 3$ is semi-stable, because for initial conditions slightly greater than 3, the solution approaches infinity, and for initial conditions slightly less than 3, the solution approaches $y = 1$.

Analysis

It is possible to find the equilibrium solutions to the differential equation by setting the right-hand side equal to zero and solving for y. This approach gives the same equilibrium solutions as those we saw in the direction field.

Exercise 4.2.2

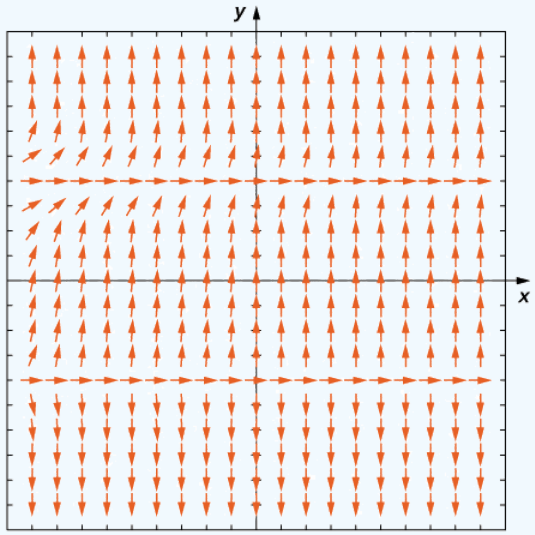
Create a direction field for the differential equation $y' = (x+5)(y+2)(y^2-4y+4)$ and identify any equilibrium solutions. Classify each of the equilibrium solutions as stable, unstable, or semi-stable.

Hint

First create the direction field and look for horizontal dashes that go all the way across. Then examine the slope lines directly above and below the equilibrium solutions.

Answer

The equilibrium solutions are $y = -2$ and $y = 2$. For this equation, $y = -2$ is an unstable equilibrium solution, and $y = 2$ is a semi-stable equilibrium solution.



Euler's Method

Consider the initial-value problem

$$y' = 2x - 3,\ y(0) = 3.$$

Integrating both sides of the differential equation gives $y = x^2 - 3x + C,$ and solving for C yields the particular solution $y = x^2 - 3x + 3$. The solution for this initial-value problem appears as the parabola in Figure 4.2.8.

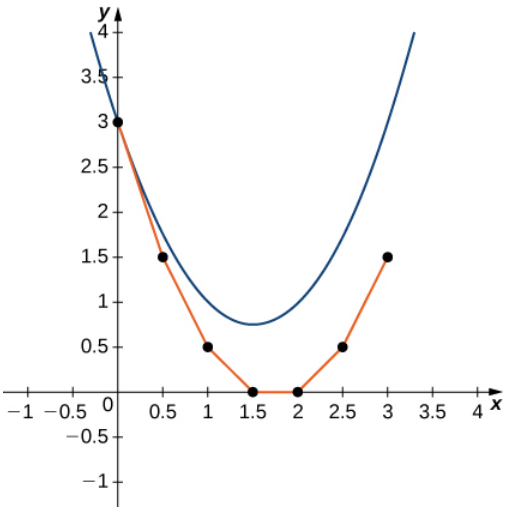


Figure 4.2.8: Euler's Method for the initial-value problem $y' = 2x - 3, y(0) = 3$.

The red graph consists of line segments that approximate the solution to the initial-value problem. The graph starts at the same initial value of $(0, 3)$. Then the slope of the solution at any point is determined by the right-hand side of the differential equation, and the length of the line segment is determined by increasing the x value by 0.5 each time (the step size). This approach is the basis of Euler's Method.

Before we state Euler's Method as a theorem, let's consider another initial-value problem:

$$y' = x^2 - y^2,\ y(-1) = 2.$$

The idea behind direction fields can also be applied to this problem to study the behavior of its solution. For example, at the point $(-1, 2)$, the slope of the solution is given by $y' = (-1)^2 - 2^2 = -3$, so the slope of the tangent line to the solution at that point is also equal to -3. Now we define $x_0 = -1$ and $y_0 = 2$. Since the slope of the solution at this point is equal to -3, we can use the method of linear approximation to approximate y near $(-1, 2)$.

$$L(x) = y_0 + f'(x_0)(x - x_0).$$

Here $x_0 = -1, y_0 = 2,$ and $f'(x_0) = -3,$ so the linear approximation becomes

$$L(x) = 2 - 3(x - (-1)) = 2 - 3x - 3 = -3x - 1.$$

Now we choose a **step size**. The step size is a small value, typically 0.1 or less, that serves as an increment for x; it is represented by the variable h. In our example, let $h = 0.1$. Incrementing x_0 by h gives our next x value:

$$x_1 = x_0 + h = -1 + 0.1 = -0.9.$$

We can substitute $x_1 = -0.9$ into the linear approximation to calculate y_1.

$$y_1 = L(x_1) = -3(-0.9) - 1 = 1.7.$$

Therefore the approximate y value for the solution when $x = -0.9$ is $y = 1.7$. We can then repeat the process, using $x_1 = -0.9$ and $y_1 = 1.7$ to calculate x_2 and y_2. The new slope is given by $y' = (-0.9)^2 - (1.7)^2 = -2.08$. First, $x_2 = x_1 + h = -0.9 + 0.1 = -0.8$. Using linear approximation gives

$$\begin{aligned} L(x) &= y_1 + f'(x_1)(x - x_1) \\ &= 1.7 - 2.08(x - (-0.9)) \\ &= 1.7 - 2.08x - 1.872 \\ &= -2.08x - 0.172. \end{aligned}$$

Finally, we substitute $x_2 = -0.8$ into the linear approximation to calculate y_2.

$$\begin{aligned} y_2 &= L(x_2) \\ &= -2.08x_2 - 0.172 \\ &= -2.08(-0.8) - 0.172 \\ &= 1.492. \end{aligned}$$

Therefore the approximate value of the solution to the differential equation is $y = 1.492$ when $x = -0.8$.

What we have just shown is the idea behind **Euler's Method**. Repeating these steps gives a list of values for the solution. These values are shown in the table, rounded off to four decimal places.

Table 4.2.1: Using Euler's Method to Approximate Solutions to a Differential Equation

n	0	1	2	3	4	5
x_n	−1	−0.9	−0.8	−0.7	−0.6	−0.5
y_n	2	1.7	1.492	1.3334	1.2046	1.0955
n	6	7	8	9	10	
x_n	−0.4	−0.3	−0.2	−0.1	0	
y_n	1.0004	1.9164	1.8414	1.7746	1.7156	

Euler's method

Consider the initial-value problem

$$y' = f(x, y),\ y(x_0) = y_0.$$

To approximate a solution to this problem using Euler's method, define

$$x_n = x_0 + nh$$
$$y_n = y_{n-1} + hf(x_{n-1}, y_{n-1})\ .$$

Here $h > 0$ represents the step size and n is an integer, starting with 1. The number of steps taken is counted by the variable n.

Typically h is a small value, say 0.1 or 0.05. The smaller the value of h, the more calculations are needed. The higher the value of h, the fewer calculations are needed. However, the tradeoff results in a lower degree of accuracy for larger step size, as illustrated in Figure $4.2.9$.

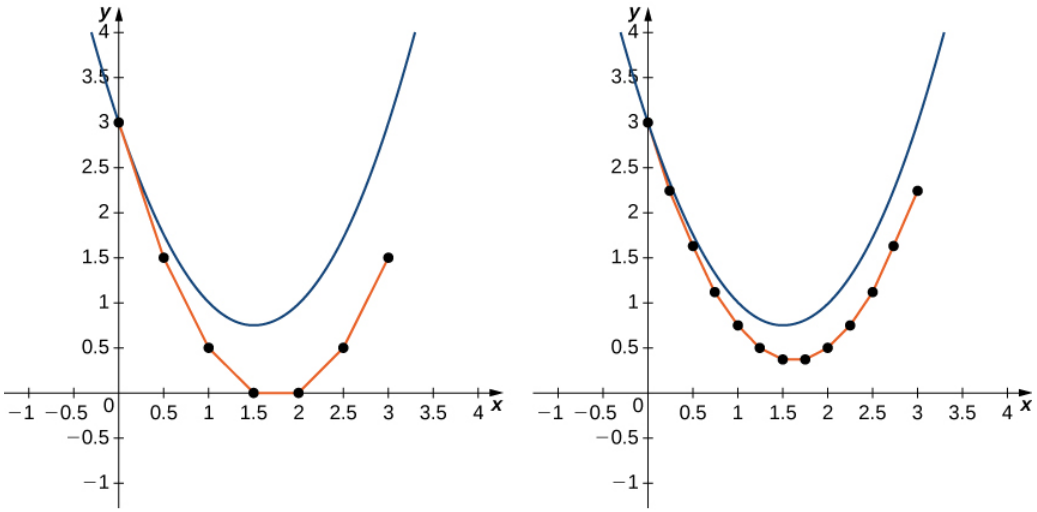


Figure 4.2.9: Euler's method for the initial-value problem $y' = 2x - 3,\ y(0) = 3$ with (a) a step size of $h = 0.5$; and (b) a step size of $h = 0.25$.

Example 4.2.2: Using Euler's Method

Consider the initial-value problem

$$y' = 3x^2 - y^2 + 1,\ y(0) = 2.$$

Use Euler's method with a step size of 0.1 to generate a table of values for the solution for values of x between 0 and 1.

Solution

We are given $h = 0.1$ and $f(x, y) = 3x^2 - y^2 + 1$. Furthermore, the initial condition $y(0) = 2$ gives $x_0 = 0$ and $y_0 = 2$. Using Equation with $n = 0$, we can generate this table.

Table 4.2.2: Using Euler's Method to Approximate Solutions to a Differential Equation

n	x_n	$y_n = y_{n-1} + hf(x_{n-1}, y_{n-1})$
0	0	2
1	0.1	$y_1 = y_0 + hf(x_0, y_0) = 1.7$
2	0.2	$y_2 = y_1 + hf(x_1, y_1) = 1.514$
3	0.3	$y_3 = y_2 + hf(x_2, y_2) = 1.3968$

n	x_n	$y_n = y_{n-1} + hf(x_{n-1}, y_{n-1})$
4	0.4	$y_4 = y_3 + hf(x_3, y_3) = 1.3287$
5	0.5	$y_5 = y_4 + hf(x_4, y_4) = 1.3001$
6	0.6	$y_6 = y_5 + hf(x_5, y_5) = 1.3061$
7	0.7	$y_7 = y_6 + hf(x_6, y_6) = 1.3435$
8	0.8	$y_8 = y_7 + hf(x_7, y_7) = 1.4100$
9	0.9	$y_9 = y_8 + hf(x_8, y_8) = 1.5032$
10	1.0	$y_{10} = y_9 + hf(x_9, y_9) = 1.6202$

With ten calculations, we are able to approximate the values of the solution to the initial-value problem for values of x between 0 and 1.

Go to this website for a chance to visually explore Euler's method.

Exercise 4.2.3

Consider the initial-value problem

$$y' = x^3 + y^2,\ y(1) = -2.$$

Using a step size of 0.1, generate a table with approximate values for the solution to the initial-value problem for values of x between 1 and 2.

Hint

Start by identifying the value of h, then figure out what $f(x, y)$ is. Then use the formula for Euler's Method to calculate y_1, y_2, and so on.

Answer

Table 4.2.3: Using Euler's Method to approximate solutions to the differential equation in Exercise 4.2.3.

n	\(x_n)	$y_n = y_{n-1} + hf(x_{n-1}, y_{n-1})$
0	1	-2
1	1.1	$y_1 = y_0 + hf(x_0, y_0) = -1.5$
2	1.2	$y_2 = y_1 + hf(x_1, y_1) = -1.1419$
3	1.3	$y_3 = y_2 + hf(x_2, y_2) = -0.8387$
4	1.4	$y_4 = y_3 + hf(x_3, y_3) = -0.5487$
5	1.5	$y_5 = y_4 + hf(x_4, y_4) = -0.2442$
6	1.6	$y_6 = y_5 + hf(x_5, y_5) = 0.0993$
7	1.7	$y_7 = y_6 + hf(x_6, y_6) = 0.5099$
8	1.8	$y_8 = y_7 + hf(x_7, y_7) = 1.0272$
9	1.9	$y_9 = y_8 + hf(x_8, y_8) = 1.7159$
10	2	$y_{10} = y_9 + hf(x_9, y_9) = 2.6962$

Key Concepts

- A direction field is a mathematical object used to graphically represent solutions to a first-order differential equation.
- Euler's Method is a numerical technique that can be used to approximate solutions to a differential equation.

Key Equations

- **Euler's Method**

  $x_n = x_0 + nh$

  $y_n = y_{n-1} + hf(x_{n-1}, y_{n-1})$, where h is the step size

Glossary

asymptotically semi-stable solution

$y = k$ if it is neither asymptotically stable nor asymptotically unstable

asymptotically stable solution

$y = k$ if there exists $\varepsilon > 0$ such that for any value $c \in (k - \varepsilon,\ k + \varepsilon)$ the solution to the initial-value problem $y' = f(x, y),\ y(x_0) = c$ approaches k as x approaches infinity

asymptotically unstable solution

$y = k$ if there exists $\varepsilon > 0$ such that for any value $c \in (k - \varepsilon,\ k + \varepsilon)$ the solution to the initial-value problem $y' = f(x, y),\ y(x_0) = c$ never approaches k as $x$approaches infinity

direction field (slope field)

a mathematical object used to graphically represent solutions to a first-order differential equation; at each point in a direction field, a line segment appears whose slope is equal to the slope of a solution to the differential equation passing through that point

equilibrium solution

any solution to the differential equation of the form $y = c$, where c is a constant

Euler's Method

a numerical technique used to approximate solutions to an initial-value problem

solution curve

a curve graphed in a direction field that corresponds to the solution to the initial-value problem passing through a given point in the direction field

step size

the increment hh that is added to the xx value at each step in Euler's Method

- Gilbert Strang (MIT) and Edwin "Jed" Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.
- Paul Seeburger (Monroe Community College) edited this section to adjust the explanation of equilibrium points in the example shown in Figures 4.2.4 - 4.2.6. He also created these figures.

This page titled 4.2: Direction Fields and Numerical Methods is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **8.2: Direction Fields and Numerical Methods** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

4.3: Separable Equations

Learning Objectives

- Use separation of variables to solve a differential equation.
- Solve applications using separation of variables.

We now examine a solution technique for finding exact solutions to a class of differential equations known as separable differential equations. These equations are common in a wide variety of disciplines, including physics, chemistry, and engineering. We illustrate a few applications at the end of the section.

Separation of Variables

We start with a definition and some examples.

Definition: Separable Differential Equations

A *separable differential equation* is any equation that can be written in the form

$$y' = f(x)g(y). \tag{4.3.1}$$

The term 'separable' refers to the fact that the right-hand side of Equation 4.3.1 can be separated into a function of x times a function of y. Examples of separable differential equations include

$$y' = (x^2-4)(3y+2) \tag{4.3.2}$$

$$y' = 6x^2+4x \tag{4.3.3}$$

$$y' = \sec y + \tan y \tag{4.3.4}$$

$$y' = xy+3x-2y-6. \tag{4.3.5}$$

Equation 4.3.3 is separable with $f(x) = 6x^2+4x$ and $g(y) = 1$, Equation 4.3.4 is separable with $f(x) = 1$ and $g(y) = \sec y + \tan y$, and the right-hand side of Equation 4.3.5 can be factored as $(x+3)(y-2)$, so it is separable as well. Equation 4.3.4 is also called an *autonomous differential equation* because the right-hand side of the equation is a function of y alone. If a differential equation is separable, then it is possible to solve the equation using the method of *separation of variables*.

Problem-Solving Strategy: Separation of Variables

1. Check for any values of y that make $g(y) = 0$. These correspond to constant solutions.
2. Rewrite the differential equation in the form
$$\frac{dy}{g(y)} = f(x)dx.$$
3. Integrate both sides of the equation.
4. Solve the resulting equation for y if possible.
5. If an initial condition exists, substitute the appropriate values for x and y into the equation and solve for the constant.

Note that Step 4 states "Solve the resulting equation for y if possible." It is not always possible to obtain y as an explicit function of x. Quite often we have to be satisfied with finding y as an implicit function of x.

Example 4.3.1: Using Separation of Variables

Find a general solution to the differential equation $y' = (x^2-4)(3y+2)$ using the method of separation of variables.

Solution

Follow the five-step method of separation of variables.

1. In this example, $f(x) = x^2 - 4$ and $g(y) = 3y + 2$. Setting $g(y) = 0$ gives $y = -\frac{2}{3}$ as a constant solution.

2. Rewrite the differential equation in the form

$$\frac{dy}{3y+2} = (x^2 - 4)\,dx.$$

3. Integrate both sides of the equation:

$$\int \frac{dy}{3y+2} = \int (x^2 - 4)\,dx.$$

Let $u = 3y + 2$. Then $du = 3\frac{dy}{dx}\,dx$, so the equation becomes

$$\frac{1}{3}\int \frac{1}{u}\,du = \frac{1}{3}x^3 - 4x + C$$

$$\frac{1}{3}\ln|u| = \frac{1}{3}x^3 - 4x + C$$

$$\frac{1}{3}\ln|3y+2| = \frac{1}{3}x^3 - 4x + C.$$

4. To solve this equation for y, first multiply both sides of the equation by 3.

$$\ln|3y+2| = x^3 - 12x + 3C$$

Now we use some logic in dealing with the constant C. Since C represents an arbitrary constant, $3C$ also represents an arbitrary constant. If we call the second arbitrary constant C_1, where $C_1 = 3C$, the equation becomes

$$\ln|3y+2| = x^3 - 12x + C_1.$$

Now exponentiate both sides of the equation (i.e., make each side of the equation the exponent for the base e).

$$e^{\ln|3y+2|} = e^{x^3 - 12x + C_1}$$

$$|3y+2| = e^{C_1}e^{x^3 - 12x}$$

Again define a new constant $C_2 = e^{C_1}$ (note that $C_2 > 0$):

$$|3y+2| = C_2 e^{x^3 - 12x}.$$

Because of the absolute value on the left side of the equation, this corresponds to two separate equations:

$$3y+2 = C_2 e^{x^3 - 12x}$$

and

$$3y+2 = -C_2 e^{x^3 - 12x}.$$

The solution to either equation can be written in the form

$$y = \frac{-2 \pm C_2 e^{x^3 - 12x}}{3}.$$

Since $C_2 > 0$, it does not matter whether we use plus or minus, so the constant can actually have either sign. Furthermore, the subscript on the constant C is entirely arbitrary, and can be dropped. Therefore the solution can be written as

$$y = \frac{-2 + Ce^{x^3 - 12x}}{3}, \text{ where } C = \pm C_2 \text{ or } C = 0.$$

Note that in writing a single general solution in this way, we are also allowing C to equal 0. This gives us the singular solution, $y = -\frac{2}{3}$, for the given differential equation. Check that this is indeed a solution of this differential equation!

5. No initial condition is imposed, so we are finished.

Exercise 4.3.1

Use the method of separation of variables to find a general solution to the differential equation

$$y' = 2xy + 3y - 4x - 6.$$

Hint

First factor the right-hand side of the equation by grouping, then use the five-step strategy of separation of variables.

Answer

$$y = 2 + Ce^{x^2+3x}$$

Example 4.3.2: Solving an Initial-Value Problem

Using the method of separation of variables, solve the initial-value problem

$$y' = (2x+3)(y^2-4), \quad y(0) = -1.$$

Solution

Follow the five-step method of separation of variables.

1. In this example, $f(x) = 2x+3$ and $g(y) = y^2 - 4$. Setting $g(y) = 0$ gives $y = \pm 2$ as constant solutions.

2. Divide both sides of the equation by $y^2 - 4$ and multiply by dx. This gives the equation

$$\frac{dy}{y^2-4} = (2x+3)\,dx.$$

3. Next integrate both sides:

$$\int \frac{1}{y^2-4}\,dy = \int (2x+3)\,dx. \tag{4.3.6}$$

To evaluate the left-hand side, use the method of partial fraction decomposition. This leads to the identity

$$\frac{1}{y^2-4} = \frac{1}{4}\left(\frac{1}{y-2} - \frac{1}{y+2}\right).$$

Then Equation 4.3.6 becomes

$$\frac{1}{4}\int \left(\frac{1}{y-2} - \frac{1}{y+2}\right) dy = \int (2x+3)\,dx$$

$$\frac{1}{4}(\ln|y-2| - \ln|y+2|) = x^2 + 3x + C.$$

Multiplying both sides of this equation by 4 and replacing $4C$ with C_1 gives

$$\ln|y-2| - \ln|y+2| = 4x^2 + 12x + C_1$$

$$\ln\left|\frac{y-2}{y+2}\right| = 4x^2 + 12x + C_1.$$

4. It is possible to solve this equation for y. First exponentiate both sides of the equation and define $C_2 = e^{C_1}$:

$$\left|\frac{y-2}{y+2}\right| = C_2 e^{4x^2+12x}.$$

Next we can remove the absolute value and let a new constant C_3 be positive, negative, or zero, i.e., $C_3 = \pm C_2$ or $C_3 = 0$.

Then multiply both sides by $y+2$.

$$y-2=C_3(y+2)e^{4x^2+12x}$$

$$y-2=C_3ye^{4x^2+12x}+2C_3e^{4x^2+12x}.$$

Now collect all terms involving y on one side of the equation, and solve for y:

$$y-C_3ye^{4x^2+12x}=2+2C_3e^{4x^2+12x}$$

$$y\left(1-C_3e^{4x^2+12x}\right)=2+2C_3e^{4x^2+12x}$$

$$y=\frac{2+2C_3e^{4x^2+12x}}{1-C_3e^{4x^2+12x}}.$$

5. To determine the value of C_3, substitute $x=0$ and $y=-1$ into the general solution. Alternatively, we can put the same values into an earlier equation, namely the equation $\frac{y-2}{y+2}=C_3e^{4x^2+12}$. This is much easier to solve for C_3:

$$\frac{y-2}{y+2}=C_3e^{4x^2+12x}$$

$$\frac{-1-2}{-1+2}=C_3e^{4(0)^2+12(0)}$$

$$C_3=-3.$$

Therefore the solution to the initial-value problem is

$$y=\frac{2-6e^{4x^2+12x}}{1+3e^{4x^2+12x}}.$$

A graph of this solution appears in Figure 4.3.1.

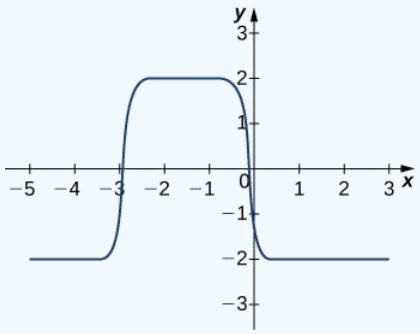


Figure 4.3.1: Graph of the solution to the initial-value problem $y'=(2x+3)(y^2-4), y(0)=-1$.

? Exercise 4.3.2

Find the solution to the initial-value problem

$$6y'=(2x+1)(y^2-2y-8)$$

with $y(0)=-3$ using the method of separation of variables.

Hint

Follow the steps for separation of variables to solve the initial-value problem.

Answer

$$y=\frac{4+14e^{x^2+x}}{1-7e^{x^2+x}}$$

Applications of Separation of Variables

Many interesting problems can be described by separable equations. We illustrate two types of problems: solution concentrations and Newton's law of cooling.

Solution concentrations

Consider a tank being filled with a salt solution. We would like to determine the amount of salt present in the tank as a function of time. We can apply the process of separation of variables to solve this problem and similar problems involving *solution concentrations*.

Example 4.3.3: Determining Salt Concentration over Time

A tank containing 100 L of a brine solution initially has 4 kg of salt dissolved in the solution. At time $t = 0$, another brine solution flows into the tank at a rate of 2 L/min. This brine solution contains a concentration of 0.5 kg/L of salt. At the same time, a stopcock is opened at the bottom of the tank, allowing the combined solution to flow out at a rate of 2 L/min, so that the level of liquid in the tank remains constant (Figure 4.3.2). Find the amount of salt in the tank as a function of time (measured in minutes), and find the limiting amount of salt in the tank, assuming that the solution in the tank is well mixed at all times.

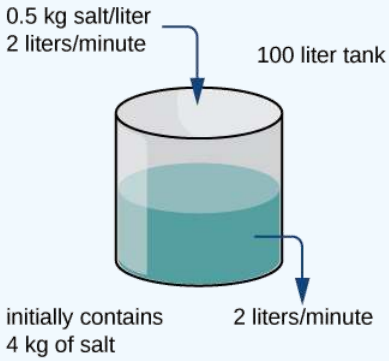


Figure 4.3.2: A brine tank with an initial amount of salt solution accepts an input flow and delivers an output flow. How does the amount of salt change with time?

Solution

First we define a function $u(t)$ that represents the amount of salt in kilograms in the tank as a function of time. Then $\frac{du}{dt}$ represents the rate at which the amount of salt in the tank changes as a function of time. Also, $u(0)$ represents the amount of salt in the tank at time $t = 0$, which is 4 kilograms.

The general setup for the differential equation we will solve is of the form

$$\frac{du}{dt} = \text{INFLOW RATE} - \text{OUTFLOW RATE}.$$

INFLOW RATE represents the rate at which salt enters the tank, and OUTFLOW RATE represents the rate at which salt leaves the tank. Because solution enters the tank at a rate of 2 L/min, and each liter of solution contains 0.5 kilogram of salt, every minute $2(0.5) = 1$ kilogram of salt enters the tank. Therefore INFLOW RATE = 1.

To calculate the rate at which salt leaves the tank, we need the concentration of salt in the tank at any point in time. Since the actual amount of salt varies over time, so does the concentration of salt. However, the volume of the solution remains fixed at 100 liters. The number of kilograms of salt in the tank at time t is equal to $u(t)$. Thus, the concentration of salt is $\frac{u(t)}{100}$ kg/L, and the solution leaves the tank at a rate of 2 L/min. Therefore salt leaves the tank at a rate of $\frac{u(t)}{100} \cdot 2 = \frac{u(t)}{50}$ kg/min, and OUTFLOW RATE is equal to $\frac{u(t)}{50}$. Therefore the differential equation becomes $\frac{du}{dt} = 1 - \frac{u}{50}$, and the initial condition is $u(0) = 4$. The initial-value problem to be solved is

$$\frac{du}{dt} = 1 - \frac{u}{50}, u(0) = 4.$$

The differential equation is a separable equation, so we can apply the five-step strategy for solution.

Step 1. Setting $1-\frac{u}{50}=0$ gives $u=50$ as a constant solution. Since the initial amount of salt in the tank is 4 kilograms, this solution does not apply.

Step 2. Rewrite the equation as

$$\frac{du}{dt}=\frac{50-u}{50}.$$

Then multiply both sides by dt and divide both sides by $50-u$:

$$\frac{du}{50-u}=\frac{dt}{50}.$$

Step 3. Integrate both sides:

$$\begin{aligned}\int\frac{du}{50-u}&=\int\frac{dt}{50}\\ -\ln|50-u|&=\frac{t}{50}+C.\end{aligned}$$

Step 4. Solve for $u(t)$:

$$\ln|50-u|=-\frac{t}{50}-C$$

$$e^{\ln|50-u|}=e^{-(t/50)-C}$$

$$|50-u|=C_1e^{-t/50},\text{ where }C_1=e^{-C}.$$

Eliminate the absolute value by allowing the constant to be positive, negative, or zero, i.e., $C_1=\pm e^{-C}$ or $C_1=0$:

$$50-u=C_1e^{-t/50}.$$

Finally, solve for $u(t)$:

$$u(t)=50-C_1e^{-t/50}.$$

Step 5. Solve for C_1:

$$\begin{aligned}u(0)&=50-C_1e^{-0/50}\\ 4&=50-C_1\\ C_1&=46.\end{aligned}$$

The solution to the initial value problem is $u(t)=50-46e^{-t/50}$. To find the limiting amount of salt in the tank, take the limit as t approaches infinity:

$$\begin{aligned}\lim_{t\to\infty}u(t)&=50-46e^{-t/50}\\ &=50-46(0)=50.\end{aligned}$$

Note that this was the constant solution to the differential equation. If the initial amount of salt in the tank is 50 kilograms, then it remains constant. If it starts at less than 50 kilograms, then it approaches 50 kilograms over time.

Exercise 4.3.3

A tank contains 3 kilograms of salt dissolved in 75 liters of water. A salt solution of 0.4 kg salt/L is pumped into the tank at a rate of 6 L/min and is drained at the same rate. Solve for the salt concentration at time t. Assume the tank is well mixed at all times.

Hint

Follow the steps in Example 4.3.3 and determine an expression for INFLOW and OUTFLOW. Formulate an initial-value problem, and then solve it.

Initial value problem:

$$\frac{du}{dt} = 2.4 - \frac{2u}{25},\ u(0) = 3$$

Answer

$$u(t) = 30 - 27e^{-t/50}$$

Newton's Law of Cooling

Newton's law of cooling states that the rate of change of an object's temperature is proportional to the difference between its own temperature and the ambient temperature (i.e., the temperature of its surroundings). If we let $T(t)$ represent the temperature of an object as a function of time, then $\frac{dT}{dt}$ represents the rate at which that temperature changes. The temperature of the object's surroundings can be represented by T_s. Then Newton's law of cooling can be written in the form

$$\frac{dT}{dt} = k(T(t) - T_s)$$

or simply

$$\frac{dT}{dt} = k(T - T_s).$$

The temperature of the object at the beginning of any experiment is the initial value for the initial-value problem. We call this temperature T_0. Therefore the initial-value problem that needs to be solved takes the form

$$\frac{dT}{dt} = k(T - T_s) \tag{4.3.7}$$

with $T(0) = T_0$, where k is a constant that needs to be either given or determined in the context of the problem. We use these equations in Example 4.3.4.

Example 4.3.4: Waiting for a Pizza to Cool

A pizza is removed from the oven after baking thoroughly, and the temperature of the oven is $350^\circ F$. The temperature of the kitchen is $75^\circ F$, and after 5 minutes the temperature of the pizza is $340^\circ F$. We would like to wait until the temperature of the pizza reaches $300^\circ F$ before cutting and serving it (Figure 4.3.3). How much longer will we have to wait?

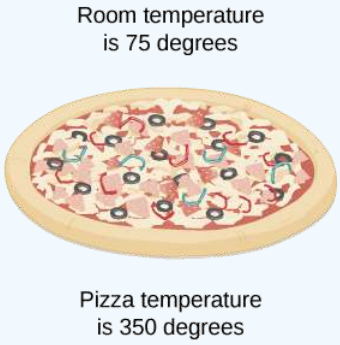


Figure 4.3.3: From Newton's law of cooling, if the pizza cools $10^\circ F$ in 5 minutes, how long before it cools to $300^\circ F$?

Solution

The ambient temperature (surrounding temperature) is $75^\circ F$, so $T_s = 75$. The temperature of the pizza when it comes out of the oven is $350^\circ F$, which is the initial temperature (i.e., initial value), so $T_0 = 350$. Therefore Equation 4.3.7 becomes

$$\frac{dT}{dt} = k(T - 75)$$

with $T(0) = 350$.

To solve the differential equation, we use the five-step technique for solving separable equations.

1. Setting the right-hand side equal to zero gives $T = 75$ as a constant solution. Since the pizza starts at $350^\circ F$, this is not the solution we are seeking.

2. Rewrite the differential equation by multiplying both sides by dt and dividing both sides by $T - 75$:

$$\frac{dT}{T-75} = k\,dt.$$

3. Integrate both sides:

$$\begin{aligned}\int \frac{dT}{T-75} &= \int k\,dt \\ \ln|T-75| &= kt + C.\end{aligned}$$

4. Solve for T by first exponentiating both sides:

$$\begin{aligned} e^{\ln|T-75|} &= e^{kt+C} \\ |T-75| &= C_1 e^{kt}, && \text{where } C_1 = e^C. \\ T-75 &= \pm C_1 e^{kt} \\ T-75 &= Ce^{kt}, && \text{where } C = \pm C_1 \text{ or } C = 0. \\ T(t) &= 75 + Ce^{kt}. \end{aligned}$$

5. Solve for C by using the initial condition $T(0) = 350$:

$$\begin{aligned} T(t) &= 75 + Ce^{kt} \\ T(0) &= 75 + Ce^{k(0)} \\ 350 &= 75 + C \\ C &= 275. \end{aligned}$$

Therefore the solution to the initial-value problem is

$$T(t) = 75 + 275e^{kt}.$$

To determine the value of k, we need to use the fact that after 5 minutes the temperature of the pizza is $340^\circ F$. Therefore $T(5) = 340$. Substituting this information into the solution to the initial-value problem, we have

$$T(t) = 75 + 275e^{kt}$$

$$T(5) = 340 = 75 + 275e^{5k}$$

$$265 = 275e^{5k}$$

$$e^{5k} = \frac{53}{55}$$

$$\ln e^{5k} = \ln\left(\frac{53}{55}\right)$$

$$5k = \ln\left(\frac{53}{55}\right)$$

$$k = \frac{1}{5}\ln\left(\frac{53}{55}\right) \approx -0.007408.$$

So now we have $T(t) = 75 + 275e^{-0.007048t}$. When is the temperature $300^\circ F$? Solving for t, we find

$$T(t) = 75 + 275e^{-0.007048t}$$

$$300 = 75 + 275e^{-0.007048t}$$

$$225 = 275e^{-0.007048t}$$

$$e^{-0.007048t} = \frac{9}{11}$$

$$\ln e^{-0.007048t} = \ln \frac{9}{11}$$

$$-0.007048t = \ln \frac{9}{11}$$

$$t = -\frac{1}{0.007048} \ln \frac{9}{11} \approx 28.5.$$

Therefore we need to wait an additional 23.5 minutes (after the temperature of the pizza reached $340^\circ F$). That should be just enough time to finish this calculation.

Exercise 4.3.4

A cake is removed from the oven after baking thoroughly, and the temperature of the oven is $450^\circ F$. The temperature of the kitchen is $70^\circ F$, and after 10 minutes the temperature of the cake is $430^\circ F$.

a. Write the appropriate initial-value problem to describe this situation.
b. Solve the initial-value problem for $T(t)$.
c. How long will it take until the temperature of the cake is within $5^\circ F$ of room temperature?

Hint

Determine the values of T_s and T_0 then use Equation 4.3.7.

Answer a

Initial-value problem

$$\frac{dT}{dt} = k(T - 70), \quad T(0) = 450$$

Answer b

$$T(t) = 70 + 380e^{kt}$$

Answer c

Approximately 114 minutes.

Key Concepts

- A separable differential equation is any equation that can be written in the form $y' = f(x)g(y)$.
- The method of separation of variables is used to find the general solution to a separable differential equation.

Key Equations

- **Separable differential equation**

  $y' = f(x)g(y)$

- **Solution concentration**

  $\frac{du}{dt} = \text{INFLOW RATE} - \text{OUTFLOW RATE}$

- **Newton's law of cooling**

  $\frac{dT}{dt} = k(T - T_s)$

Glossary

autonomous differential equation

an equation in which the right-hand side is a function of y alone

separable differential equation

any equation that can be written in the form $y' = f(x)g(y)$

separation of variables

a method used to solve a separable differential equation

This page titled 4.3: Separable Equations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **8.3: Separable Equations** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

4.4: The Logistic Equation

Learning Objectives

- Describe the concept of environmental carrying capacity in the logistic model of population growth.
- Draw a direction field for a logistic equation and interpret the solution curves.
- Solve a logistic equation and interpret the results.

Differential equations can be used to represent the size of a population as it varies over time. We saw this in an earlier chapter in the section on exponential growth and decay, which is the simplest model. A more realistic model includes other factors that affect the growth of the population. In this section, we study the logistic differential equation and see how it applies to the study of population dynamics in the context of biology.

Population Growth and Carrying Capacity

To model population growth using a differential equation, we first need to introduce some variables and relevant terms. The variable t. will represent time. The units of time can be hours, days, weeks, months, or even years. Any given problem must specify the units used in that particular problem. The variable P will represent population. Since the population varies over time, it is understood to be a function of time. Therefore we use the notation $P(t)$ for the population as a function of time. If $P(t)$ is a differentiable function, then the first derivative $\frac{dP}{dt}$ represents the instantaneous rate of change of the population as a function of time.

In Exponential Growth and Decay, we studied the exponential growth and decay of populations and radioactive substances. An example of an exponential growth function is $P(t) = P_0 e^{rt}$. In this function, $P(t)$ represents the population at time t, P_0 represents the **initial population** (population at time $t = 0$), and the constant $r > 0$ is called the **growth rate**. Figure 4.4.1 shows a graph of $P(t) = 100e^{0.03t}$. Here $P_0 = 100$ and $r = 0.03$.

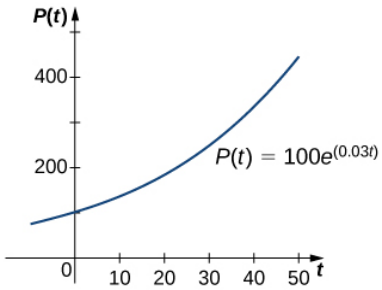


Figure 4.4.1: An exponential growth model of population.

We can verify that the function $P(t) = P_0 e^{rt}$ satisfies the initial-value problem

$$\frac{dP}{dt} = rP$$

with $P(0) = P_0$.

This differential equation has an interesting interpretation. The left-hand side represents the rate at which the population increases (or decreases). The right-hand side is equal to a positive constant multiplied by the current population. Therefore the differential equation states that the rate at which the population increases is proportional to the population at that point in time. Furthermore, it states that the constant of proportionality never changes.

One problem with this function is its prediction that as time goes on, the population grows without bound. This is unrealistic in a real-world setting. Various factors limit the rate of growth of a particular population, including birth rate, death rate, food supply, predators, and so on. The growth constant r usually takes into consideration the birth and death rates but none of the other factors, and it can be interpreted as a net (birth minus death) percent growth rate per unit time. A natural question to ask is whether the population growth rate stays constant, or whether it changes over time. Biologists have found that in many biological systems, the population grows until a certain steady-state population is reached. This possibility is not taken into account with exponential growth. However, the concept of carrying capacity allows for the possibility that in a given area, only a certain number of a given organism or animal can thrive without running into resource issues.

Definition: Carrying Capacity

The *carrying capacity* of an organism in a given environment is defined to be the maximum population of that organism that the environment can sustain indefinitely.

We use the variable K to denote the carrying capacity. The growth rate is represented by the variable r. Using these variables, we can define the logistic differential equation.

Definition: Logistic Differential Equation

Let K represent the carrying capacity for a particular organism in a given environment, and let r be a real number that represents the growth rate. The function $P(t)$ represents the population of this organism as a function of time t, and the constant P_0 represents the initial population (population of the organism at time $t = 0$). Then the *logistic differential equation* is

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right). \tag{4.4.1}$$

The logistic equation was first published by Pierre Verhulst in 1845. This differential equation can be coupled with the initial condition $P(0) = P_0$ to form an initial-value problem for $P(t)$.

Suppose that the initial population is small relative to the carrying capacity. Then $\frac{P}{K}$ is small, possibly close to zero. Thus, the quantity in parentheses on the right-hand side of Equation 4.4.1 is close to 1, and the right-hand side of this equation is close to rP. If $r > 0$, then the population grows rapidly, resembling exponential growth.

However, as the population grows, the ratio $\frac{P}{K}$ also grows, because K is constant. If the population remains below the carrying capacity, then $\frac{P}{K}$ is less than 1, so $1 - \frac{P}{K} > 0$. Therefore the right-hand side of Equation 4.4.1 is still positive, but the quantity in parentheses gets smaller, and the growth rate decreases as a result. If $P = K$ then the right-hand side is equal to zero, and the population does not change.

Now suppose that the population starts at a value higher than the carrying capacity. Then $\frac{P}{K} > 1$, and $1 - \frac{P}{K} < 0$. Then the right-hand side of Equation 4.4.1 is negative, and the population decreases. As long as $P > K$, the population decreases. It never actually reaches K because $\frac{dP}{dt}$ will get smaller and smaller, but the population approaches the carrying capacity as t approaches infinity. This analysis can be represented visually by way of a phase line. A **phase line** describes the general behavior of a solution to an autonomous differential equation, depending on the initial condition. For the case of a carrying capacity in the logistic equation, the phase line is as shown in Figure 4.4.2.



Figure 4.4.2: A phase line for the differential equation $\frac{dP}{dt} = rP(1 - \frac{P}{K})$.

This phase line shows that when P is less than zero or greater than K, the population decreases over time. When P is between 0 and K, the population increases over time.

Example 4.4.1: Examining the Carrying Capacity of a Deer Population

Let's consider the population of white-tailed deer (*Odocoileus virginianus*) in the state of Kentucky. The Kentucky Department of Fish and Wildlife Resources (KDFWR) sets guidelines for hunting and fishing in the state. Before the hunting season of 2004, it estimated a population of 900,000 deer. Johnson notes: "A deer population that has plenty to eat and is not hunted by humans or other predators will double every three years." (George Johnson, "The Problem of Exploding Deer Populations Has No Attractive Solutions," January 12,2001, accessed April 9, 2015)



Figure 4.4.3: (credit: modification of work by Rachel Kramer, Flickr)

This observation corresponds to a rate of increase $r = \dfrac{\ln(2)}{3} = 0.2311$, so the approximate growth rate is 23.11% per year. (This assumes that the population grows exponentially, which is reasonable—at least in the short term—with plentiful food supply and no predators.) The KDFWR also reports deer population densities for 32 counties in Kentucky, the average of which is approximately 27 deer per square mile. Suppose this is the deer density for the whole state (39,732 square miles). The carrying capacity K is 39,732 square miles times 27 deer per square mile, or 1,072,764 deer.

a. For this application, we have $P_0 = 900,000$, $K = 1,072,764$, and $r = 0.2311$. Substitute these values into Equation 4.4.1 and form the initial-value problem.
b. Solve the initial-value problem from part a.
c. According to this model, what will be the population in 3 years? Recall that the doubling time predicted by Johnson for the deer population was 3 years. How do these values compare?

Suppose the population managed to reach **1,200,000** What does the logistic equation predict will happen to the population in this scenario?

Solution

a. The initial value problem is

$$\frac{dP}{dt} = 0.2311P\left(1 - \frac{P}{1,072,764}\right), \quad P(0) = 900,000.$$

b. The logistic equation is an autonomous differential equation, so we can use the method of separation of variables.

Step 1: Setting the right-hand side equal to zero gives $P = 0$ and $P = 1,072,764$. This means that if the population starts at zero it will never change, and if it starts at the carrying capacity, it will never change.

Step 2: Rewrite the differential equation and multiply both sides by:

$$\frac{dP}{dt} = 0.2311P\left(\frac{1,072,764 - P}{1,072,764}\right)$$

$$dP = 0.2311P\left(\frac{1,072,764 - P}{1,072,764}\right)dt$$

$$\frac{dP}{P(1,072,764 - P)} = \frac{0.2311}{1,072,764}dt.$$

Step 3: Integrate both sides of the equation using partial fraction decomposition:

$$\int \frac{dP}{P(1,072,764-P)} = \int \frac{0.2311}{1,072,764} dt$$

$$\frac{1}{1,072,764} \int \left(\frac{1}{P} + \frac{1}{1,072,764-P} \right) dP = \frac{0.2311t}{1,072,764} + C$$

$$\frac{1}{1,072,764} (\ln|P| - \ln|1,072,764-P|) = \frac{0.2311t}{1,072,764} + C.$$

Step 4: Multiply both sides by 1,072,764 and use the quotient rule for logarithms:

$$\ln\left|\frac{P}{1,072,764-P}\right| = 0.2311t + C_1.$$

Here $C_1 = 1,072,764C$. Next exponentiate both sides and eliminate the absolute value:

$$e^{\ln\left|\frac{P}{1,072,764-P}\right|} = e^{0.2311t+C_1}$$

$$\left|\frac{P}{1,072,764-P}\right| = C_2 e^{0.2311t}$$

$$\frac{P}{1,072,764-P} = C_2 e^{0.2311t}.$$

Here $C_2 = e^{C_1}$ but after eliminating the absolute value, it can be negative as well. Now solve for:

$$P = C_2 e^{0.2311t}(1,072,764-P)$$

$$P = 1,072,764 C_2 e^{0.2311t} - C_2 P e^{0.2311t}$$

$$P + C_2 P e^{0.2311t} = 1,072,764 C_2 e^{0.2311t}$$

$$P(1 + C_2 e^{0.2311t} = 1,072,764 C_2 e^{0.2311t}$$

$$P(t) = \frac{1,072,764 C_2 e^{0.2311t}}{1 + C_2 e^{0.2311t}}.$$

Step 5: To determine the value of C_2, it is actually easier to go back a couple of steps to where C_2 was defined. In particular, use the equation

$$\frac{P}{1,072,764-P} = C_2 e^{0.2311t}.$$

The initial condition is $P(0) = 900,000$. Replace P with $900,000$ and t with zero:

$$\frac{P}{1,072,764-P} = C_2 e^{0.2311t}$$

$$\frac{900,000}{1,072,764-900,000} = C_2 e^{0.2311(0)}$$

$$\frac{900,000}{172,764} = C_2$$

$$C_2 = \frac{25,000}{4,799}$$

$$\approx 5.209.$$

Therefore

$$P(t) = \frac{1,072,764\left(\frac{25000}{4799}\right)e^{0.2311t}}{1+(250004799)e^{0.2311t}}$$
$$= \frac{1,072,764(25000)e^{0.2311t}}{4799+25000e^{0.2311t}}.$$

Dividing the numerator and denominator by **25,000** gives

$$P(t) = \frac{1,072,764e^{0.2311t}}{0.19196+e^{0.2311t}}.$$

Figure is a graph of this equation.

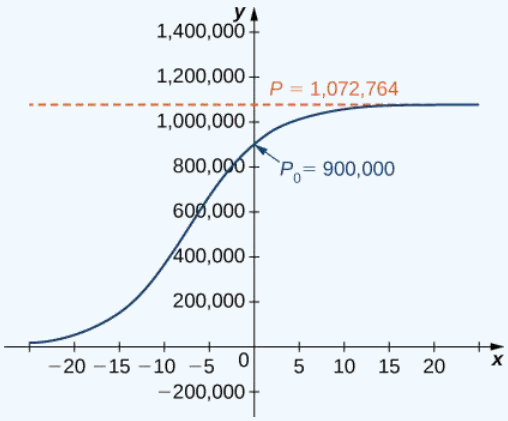


Figure 4.4.4: Logistic curve for the deer population with an initial population of **900,000** deer.

c. Using this model we can predict the population in **3** years.

$$P(3) = \frac{1,072,764e^{0.2311(3)}}{0.19196+e^{0.2311(3)}} \approx 978,830\, deer$$

This is far short of twice the initial population of $900,000$. Remember that the doubling time is based on the assumption that the growth rate never changes, but the logistic model takes this possibility into account.

d. If the population reached **1,200,000** deer, then the new initial-value problem would be

$$\frac{dP}{dt} = 0.2311P\left(1-\frac{P}{1,072,764}\right),\ P(0) = 1,200,000.$$

The general solution to the differential equation would remain the same.

$$P(t) = \frac{1,072,764C_2e^{0.2311t}}{1+C_2e^{0.2311t}}$$

To determine the value of the constant, return to the equation

$$\frac{P}{1,072,764-P} = C_2e^{0.2311t}.$$

Substituting the values $t = 0$ and $P = 1,200,000$, you get

$$C_2e^{0.2311(0)} = \frac{1,200,000}{1,072,764-1,200,000}$$
$$C_2 = -\frac{100,000}{10,603} \approx -9.431.$$

Therefore

$$P(t) = \frac{1,072,764C_2e^{0.2311t}}{1+C_2e^{0.2311t}}$$

$$= \frac{1,072,764\left(-\frac{100,000}{10,603}\right)e^{0.2311t}}{1+\left(-\frac{100,000}{10,603}\right)e^{0.2311t}}$$

$$= -\frac{107,276,400,000e^{0.2311t}}{100,000e^{0.2311t}-10,603}$$

$$\approx \frac{10,117,551e^{0.2311t}}{9.43129e^{0.2311t}-1}$$

This equation is graphed in Figure 4.4.5.

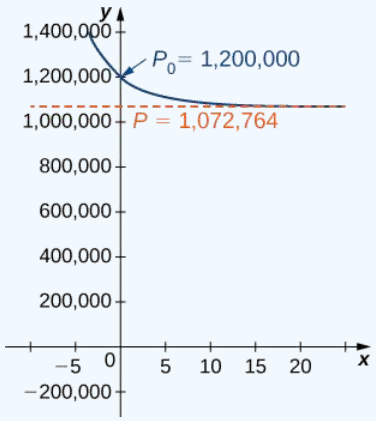


Figure 4.4.5: Logistic curve for the deer population with an initial population of **1,200,000** deer.

Solving the Logistic Differential Equation

The logistic differential equation is an autonomous differential equation, so we can use separation of variables to find the general solution, as we just did in Example 4.4.1.

Step 1: Setting the right-hand side equal to zero leads to $P = 0$ and $P = K$ as constant solutions. The first solution indicates that when there are no organisms present, the population will never grow. The second solution indicates that when the population starts at the carrying capacity, it will never change.

Step 2: Rewrite the differential equation in the form

$$\frac{dP}{dt} = \frac{rP(K-P)}{K}.$$

Then multiply both sides by dt and divide both sides by $P(K-P)$. This leads to

$$\frac{dP}{P(K-P)} = \frac{r}{K}dt.$$

Multiply both sides of the equation by K and integrate:

$$\int \frac{K}{P(K-P)}dP = \int rdt. \tag{4.4.2}$$

The left-hand side of this equation can be integrated using partial fraction decomposition. We leave it to you to verify that

$$\frac{K}{P(K-P)} = \frac{1}{P} + \frac{1}{K-P}.$$

Then the Equation 4.4.2 becomes

$$\int \frac{1}{P} + \frac{1}{K-P} dP = \int r dt$$

$$\ln|P| - \ln|K-P| = rt + C$$

$$\ln\left|\frac{P}{K-P}\right| = rt + C.$$

Now exponentiate both sides of the equation to eliminate the natural logarithm:

$$e^{\ln\left|\frac{P}{K-P}\right|} = e^{rt+C}$$

$$\left|\frac{P}{K-P}\right| = e^{C} e^{rt}.$$

We define $C_1 = e^c$ so that the equation becomes

$$\frac{P}{K-P} = C_1 e^{rt}. \tag{4.4.3}$$

To solve this equation for $P(t)$, first multiply both sides by $K-P$ and collect the terms containing P on the left-hand side of the equation:

$$P = C_1 e^{rt}(K-P)$$

$$= C_1 K e^{rt} - C_1 P e^{rt}$$

$$P + C_1 P e^{rt} = C_1 K e^{rt}.$$

Next, factor P from the left-hand side and divide both sides by the other factor:

$$P(1 + C_1 e^{rt}) = C_1 K e^{rt}$$

$$P(t) = \frac{C_1 K e^{rt}}{1 + C_1 e^{rt}}.$$

The last step is to determine the value of C_1. The easiest way to do this is to substitute $t = 0$ and P_0 in place of P in Equation and solve for C_1:

$$\frac{P}{K-P} = C_1 e^{rt}$$

$$\frac{P_0}{K-P_0} = C_1 e^{r(0)}$$

$$C_1 = \frac{P_0}{K-P_0}.$$

Finally, substitute the expression for C_1 into Equation 4.4.3:

$$P(t) = \frac{C_1 K e^{rt}}{1 + C_1 e^{rt}} = \frac{\frac{P_0}{K-P_0} K e^{rt}}{1 + \frac{P_0}{K-P_0} e^{rt}}$$

Now multiply the numerator and denominator of the right-hand side by $(K-P_0)$ and simplify:

$$P(t) = \frac{\frac{P_0}{K-P_0}Ke^{rt}}{1+\frac{P_0}{K-P_0}e^{rt}}$$

$$= \frac{\frac{P_0}{K-P_0}Ke^{rt}}{1+\frac{P_0}{K-P_0}e^{rt}} \cdot \frac{K-P_0}{K-P_0} = \frac{P_0Ke^{rt}}{(K-P_0)+P_0e^{rt}}.$$

We state this result as a theorem.

Solution of the Logistic Differential Equation

Consider the logistic differential equation subject to an initial population of P_0 with carrying capacity K and growth rate r. The solution to the corresponding initial-value problem is given by

$$P(t) = \frac{P_0Ke^{rt}}{(K-P_0)+P_0e^{rt}}$$

.

Now that we have the solution to the initial-value problem, we can choose values for P_0, r, and K and study the solution curve. For example, in Example we used the values $r = 0.2311, K = 1,072,764,$ and an initial population of $900,000$ deer. This leads to the solution

$$P(t) = \frac{P_0Ke^{rt}}{(K-P_0)+P_0e^{rt}}$$

$$= \frac{900,000(1,072,764)e^{0.2311t}}{(1,072,764-900,000)+900,000e^{0.2311t}}$$

$$= \frac{900,000(1,072,764)e^{0.2311t}}{172,764+900,000e^{0.2311t}}.$$

Dividing top and bottom by $900,000$ gives

$$P(t) = \frac{1,072,764e^{0.2311t}}{0.19196+e^{0.2311t}}.$$

This is the same as the original solution. The graph of this solution is shown again in blue in Figure 4.4.6, superimposed over the graph of the exponential growth model with initial population $900,000$ and growth rate 0.2311 (appearing in green). The red dashed line represents the carrying capacity, and is a horizontal asymptote for the solution to the logistic equation.

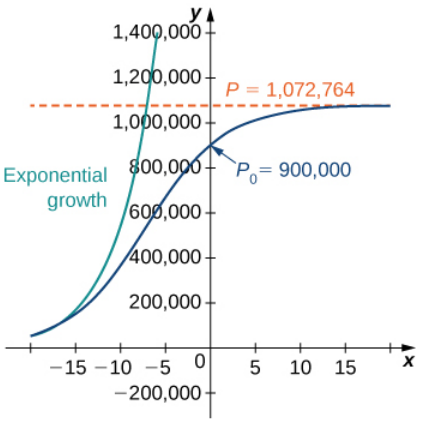


Figure 4.4.6: A comparison of exponential versus logistic growth for the same initial population of $900,000$ organisms and growth rate of 23.11

Working under the assumption that the population grows according to the logistic differential equation, this graph predicts that approximately 20 years earlier (1984), the growth of the population was very close to exponential. The net growth rate at that time would have been around 23.1 per year. As time goes on, the two graphs separate. This happens because the population increases, and the logistic differential equation states that the growth rate decreases as the population increases. At the time the population was measured (2004), it was close to carrying capacity, and the population was starting to level off.

The solution to the logistic differential equation has a point of inflection. To find this point, set the second derivative equal to zero:

$$P(t) = \frac{P_0 K e^{rt}}{(K - P_0) + P_0 e^{rt}}$$

$$P'(t) = \frac{rP_0 K(K - P0)e^{rt}}{((K - P_0) + P_0 e^{rt})^2}$$

$$\begin{aligned} P''(t) &= \frac{r^2 P_0 K(K - P_0)^2 e^{rt} - r^2 P_0^2 K(K - P_0) e^{2rt}}{((K - P_0) + P_0 e^{rt})^3} \\ &= \frac{r^2 P_0 K(K - P_0) e^{rt}((K - P_0) - P_0 e^{rt})}{((K - P_0) + P_0 e^{rt})^3}. \end{aligned}$$

Setting the numerator equal to zero,

$$r^2 P_0 K(K - P_0) e^{rt}((K - P_0) - P_0 e^{rt}) = 0.$$

As long as $P_0 \neq K$, the entire quantity before and including $e^{rt}$is nonzero, so we can divide it out:

$$(K - P_0) - P_0 e^{rt} = 0.$$

Solving for t,

$$P_0 e^{rt} = K - P_0$$

$$e^{rt} = \frac{K - P_0}{P_0}$$

$$\ln e^{rt} = \ln \frac{K - P_0}{P_0}$$

$$rt = \ln \frac{K - P_0}{P_0}$$

$$t = \frac{1}{r} \ln \frac{K - P_0}{P_0}.$$

Notice that if $P_0 > K$, then this quantity is undefined, and the graph does not have a point of inflection. In the logistic graph, the point of inflection can be seen as the point where the graph changes from concave up to concave down. This is where the “leveling off” starts to occur, because the net growth rate becomes slower as the population starts to approach the carrying capacity.

? Exercise 4.4.1

A population of rabbits in a meadow is observed to be 200 rabbits at time $t = 0$. After a month, the rabbit population is observed to have increased by 4. Using an initial population of 200 and a growth rate of 0.04, with a carrying capacity of 750 rabbits,

a. Write the logistic differential equation and initial condition for this model.
b. Draw a slope field for this logistic differential equation, and sketch the solution corresponding to an initial population of 200 rabbits.
c. Solve the initial-value problem for $P(t)$.
d. Use the solution to predict the population after 1 year.

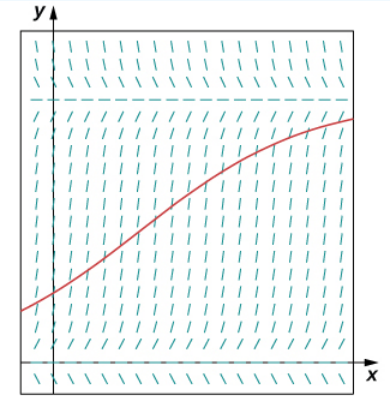
Hint

First determine the values of $r, K,$ and P_0. Then create the initial-value problem, draw the direction field, and solve the problem.

Answer

a. $\frac{dP}{dt} = 0.04(1 - \frac{P}{750}), P(0) = 200$

b.



c. $P(t) = \frac{3000e^{.04t}}{11 + 4e^{.04t}}$

d. After 12 months, the population will be $P(12) \approx 278$ rabbits.

Student Project: Logistic Equation with a Threshold Population

An improvement to the logistic model includes a **threshold population.** The threshold population is defined to be the minimum population that is necessary for the species to survive. We use the variable T to represent the threshold population. A differential equation that incorporates both the threshold population T and carrying capacity K is

$$\frac{dP}{dt} = -rP\left(1 - \frac{P}{K}\right)\left(1 - \frac{P}{T}\right)$$

where r represents the growth rate, as before.

1. The threshold population is useful to biologists and can be utilized to determine whether a given species should be placed on the endangered list. A group of Australian researchers say they have determined the threshold population for any species to survive: 5000 adults. (Catherine Clabby, “A Magic Number,” American Scientist 98(1): 24, doi:10.1511/2010.82.24. accessed April 9, 2015, www.americanscientist.org/iss...a-magic-number). Therefore we use $T = 5000$ as the threshold population in this project. Suppose that the environmental carrying capacity in Montana for elk is $25,000$. Set up Equation using the carrying capacity of $25,000$ and threshold population of 5000. Assume an annual net growth rate of 18%.
2. Draw the direction field for the differential equation from step 1, along with several solutions for different initial populations. What are the constant solutions of the differential equation? What do these solutions correspond to in the original population model (i.e., in a biological context)?
3. What is the limiting population for each initial population you chose in step 2? (Hint: use the slope field to see what happens for various initial populations, i.e., look for the horizontal asymptotes of your solutions.)
4. This equation can be solved using the method of separation of variables. However, it is very difficult to get the solution as an explicit function of t. Using an initial population of $18,000$ elk, solve the initial-value problem and express the solution as an implicit function of t, or solve the general initial-value problem, finding a solution in terms of $r, K, T,$ and P_0.

Key Concepts

- When studying population functions, different assumptions—such as exponential growth, logistic growth, or threshold population—lead to different rates of growth.

- The logistic differential equation incorporates the concept of a carrying capacity. This value is a limiting value on the population for any given environment.
- The logistic differential equation can be solved for any positive growth rate, initial population, and carrying capacity.

Key Equations

- **Logistic differential equation and initial-value problem**

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right), \quad P(0) = P_0$$

- **Solution to the logistic differential equation/initial-value problem**

$$P(t) = \frac{P_0 K e^{rt}}{(K - P_0) + P_0 e^{rt}}$$

- **Threshold population model**

$$\frac{dP}{dt} = -rP\left(1 - \frac{P}{K}\right)\left(1 - \frac{P}{T}\right)$$

Glossary

carrying capacity
the maximum population of an organism that the environment can sustain indefinitely

growth rate
the constant $r > 0$ in the exponential growth function $P(t) = P_0 e^{rt}$

initial population
the population at time $t = 0$

logistic differential equation
a differential equation that incorporates the carrying capacity K and growth rate rr into a population model

phase line
a visual representation of the behavior of solutions to an autonomous differential equation subject to various initial conditions

threshold population
the minimum population that is necessary for a species to survive

---

This page titled 4.4: The Logistic Equation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **8.4: The Logistic Equation** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

CHAPTER OVERVIEW

5: Sequences and Series

The topic of infinite series may seem unrelated to differential and integral calculus. In fact, an infinite series whose terms involve powers of a variable is a powerful tool that we can use to express functions as “infinite polynomials.” We can use infinite series to evaluate complicated functions, approximate definite integrals, and create new functions. In addition, infinite series are used to solve differential equations that model physical behavior, from tiny electronic circuits to Earth-orbiting satellites.

5.E: Sequences and Series (Exercises)
5.0: Prelude to Sequence and Series
5.1: Sequences
5.2: Infinite Series
5.3: The Divergence and Integral Tests
5.4: Comparison Tests
5.5: Alternating Series
5.6: Ratio and Root Tests

Thumbnail: For the alternating harmonic series, the odd terms S_{2k+1} in the sequence of partial sums are decreasing and bounded below. The even terms S_{2k} are increasing and bounded above.

Contributors

Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 5: Sequences and Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

5.E: Sequences and Series (Exercises)

This page titled 5.E: Sequences and Series (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

5.0: Prelude to Sequence and Series

The Koch snowflake is constructed from an infinite number of nonoverlapping equilateral triangles. Consequently, we can express its area as a sum of infinitely many terms. How do we add an infinite number of terms? Can a sum of an infinite number of terms be finite? To answer these questions, we need to introduce the concept of an infinite series, a sum with infinitely many terms. Having defined the necessary tools, we will be able to calculate the area of the Koch snowflake.

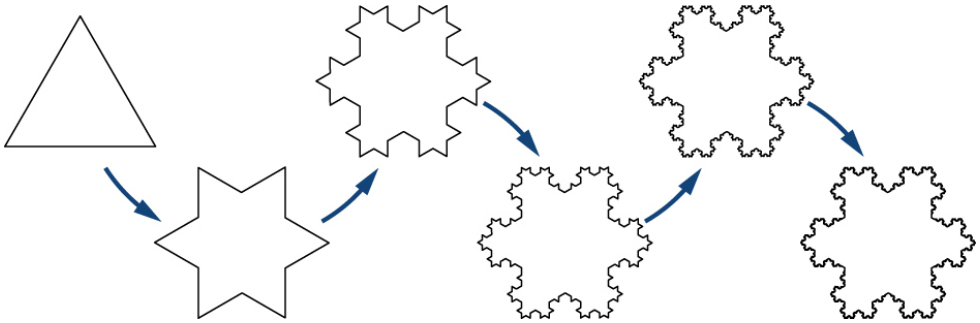


Figure 5.0.1: The Koch snowflake is constructed by using an iterative process. Starting with an equilateral triangle, at each step of the process the middle third of each line segment is removed and replaced with an equilateral triangle pointing outward.

The topic of infinite series may seem unrelated to differential and integral calculus. In fact, an infinite series whose terms involve powers of a variable is a powerful tool that we can use to express functions as “infinite polynomials.” We can use infinite series to evaluate complicated functions, approximate definite integrals, and create new functions. In addition, infinite series are used to solve differential equations that model physical behavior, from tiny electronic circuits to Earth-orbiting satellites.

This page titled 5.0: Prelude to Sequence and Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.0: Prelude to Sequence and Series** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

5.1: Sequences

Learning Objectives

- Find the formula for the general term of a sequence.
- Calculate the limit of a sequence if it exists.
- Determine the convergence or divergence of a given sequence.

In this section, we introduce sequences and define what it means for a sequence to converge or diverge. We show how to find limits of sequences that converge, often by using the properties of limits for functions discussed earlier. We close this section with the Monotone Convergence Theorem, a tool we can use to prove that certain types of sequences converge.

Terminology of Sequences

To work with this new topic, we need some new terms and definitions. First, an infinite sequence is an ordered list of numbers of the form

$$a_1, a_2, a_3, \ldots, a_n, \ldots.$$

Each of the numbers in the sequence is called a term. The symbol n is called the index variable for the sequence. We use the notation

$$\{a_n\}_{n=1}^{\infty},$$

or simply $\{a_n\}$, to denote this sequence. A similar notation is used for sets, but a sequence is an ordered list, whereas a set is not ordered. Because a particular number a_n exists for each positive integer n, we can also define a sequence as a function whose domain is the set of positive integers.

Let's consider the infinite, ordered list

$$2, 4, 8, 16, 32, \ldots.$$

This is a sequence in which the first, second, and third terms are given by $a_1 = 2, a_2 = 4,$ and $a_3 = 8.$ You can probably see that the terms in this sequence have the following pattern:

$$a_1 = 2^1,\ a_2 = 2^2,\ a_3 = 2^3,\ a_4 = 2^4 \text{ and } a_5 = 2^5.$$

Assuming this pattern continues, we can write the nth term in the sequence by the explicit formula $a_n = 2^n.$ Using this notation, we can write this sequence as

$$\{2^n\}_{n=1}^{\infty}$$

or

$$\{2^n\}.$$

Alternatively, we can describe this sequence in a different way. Since each term is twice the previous term, this sequence can be defined recursively by expressing the nth term a_n in terms of the previous term a_{n-1}. In particular, we can define this sequence as the sequence $\{a_n\}$ where $a_1 = 2$ and for all $n \geq 2$, each term an is defined by the **recurrence relation**

$$a_n = 2a_{n-1}.$$

Definition: Infinite Sequence

An **infinite sequence** $\{a_n\}$ is an ordered list of numbers of the form

$$a_1,\ a_2,\ \ldots,\ a_n,\ \ldots.$$

The subscript n is called the **index variable** of the sequence. Each number a_n is a term of the sequence. Sometimes sequences are defined by **explicit formulas,** in which case $a_n = f(n)$ for some function $f(n)$ defined over the positive integers. In other cases, sequences are defined by using a **recurrence relation**. In a recurrence relation, one term (or more) of the sequence is given explicitly, and subsequent terms are defined in terms of earlier terms in the sequence.

Note that the index does not have to start at $n = 1$ but could start with other integers. For example, a sequence given by the explicit formula $a_n = f(n)$ could start at $n = 0$, in which case the sequence would be

$$a_0,\ a_1,\ a_2, \ldots .$$

Similarly, for a sequence defined by a recurrence relation, the term a_0 may be given explicitly, and the terms a_n for $n \geq 1$ may be defined in terms of a_{n-1}. Since a sequence $\{a_n\}$ has exactly one value for each positive integer n, it can be described as a function whose domain is the set of positive integers. As a result, it makes sense to discuss the graph of a sequence. The graph of a sequence $\{a_n\}$ consists of all points (n, a_n) for all positive integers n. Figure 5.1.1 shows the graph of 2^n.

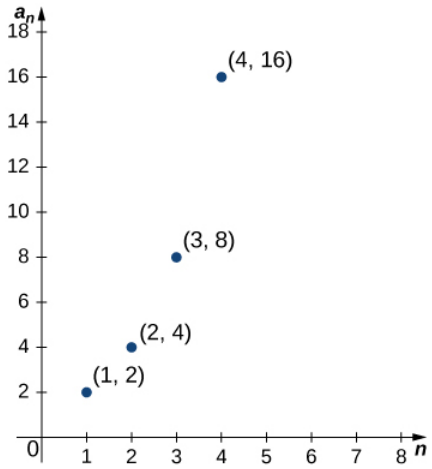


Figure 5.1.1: The plotted points are a graph of the sequence $\{2^n\}$.

Two types of sequences occur often and are given special names: arithmetic sequences and geometric sequences. In an arithmetic sequence, the difference between every pair of consecutive terms is the same. For example, consider the sequence

$$3,\ 7,\ 11,\ 15, 1\,9, \ldots$$

You can see that the difference between every consecutive pair of terms is 4. Assuming that this pattern continues, this sequence is an arithmetic sequence. It can be described by using the recurrence relation

$$\begin{cases} a_1 = 3 \\ a_n = a_{n-1} + 4, \text{ for } n \geq 2 \end{cases}.$$

Note that

$$a_2 = 3 + 4$$

$$a_3 = 3 + 4 + 4 = 3 + 2 \cdot 4$$

$$a_4 = 3 + 4 + 4 + 4 = 3 + 3 \cdot 4.$$

Thus the sequence can also be described using the explicit formula

$$a_n = 3 + 4(n-1) = 4n - 1.$$

In general, an arithmetic sequence is any sequence of the form $a_n = cn + b$.

In a geometric sequence, the ratio of every pair of consecutive terms is the same. For example, consider the sequence

$$2,\ -\frac{2}{3},\ \frac{2}{9},\ -\frac{2}{27},\ \frac{2}{81}, \ldots .$$

We see that the ratio of any term to the preceding term is $-\frac{1}{3}$. Assuming this pattern continues, this sequence is a geometric sequence. It can be defined recursively as

$$a_1 = 2$$

$$a_n = -\frac{1}{3} \cdot a_{n-1}, \text{ for } n \geq 2.$$

Alternatively, since

$$\begin{aligned} a_2 &= -\frac{1}{3} \cdot 2 \\ a_3 &= \left(-\frac{1}{3}\right)\left(-\frac{1}{3}\right)(2) = \left(-\frac{1}{3}\right)^2 \cdot 2 \\ a_4 &= \left(-\frac{1}{3}\right)\left(-\frac{1}{3}\right)\left(-\frac{1}{3}\right)(2) = \left(-\frac{1}{3}\right)^3 \cdot 2, \end{aligned}$$

we see that the sequence can be described by using the explicit formula

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

The sequence $\{2^n\}$ that we discussed earlier is a geometric sequence, where the ratio of any term to the previous term is 2. In general, a geometric sequence is any sequence of the form $a_n = cr^n$.

Example 5.1.1: Finding Explicit Formulas

For each of the following sequences, find an explicit formula for the nth term of the sequence.

a. $-\frac{1}{2}, \frac{2}{3}, -\frac{3}{4}, \frac{4}{5}, -\frac{5}{6}, \ldots$
b. $\frac{3}{4}, \frac{9}{7}, \frac{27}{10}, \frac{81}{13}, \frac{243}{16}, \ldots$.

Solution

a. First, note that the sequence is alternating from negative to positive. The odd terms in the sequence are negative, and the even terms are positive. Therefore, the nth term includes a factor of $(-1)^n$. Next, consider the sequence of numerators $1, 2, 3, \ldots$ and the sequence of denominators $2, 3, 4, \ldots$. We can see that both of these sequences are arithmetic sequences. The nth term in the sequence of numerators is n, and the nth term in the sequence of denominators is $n+1$. Therefore, the sequence can be described by the explicit formula

$$a_n = \frac{(-1)^n n}{n+1}.$$

b. The sequence of numerators $3, 9, 27, 81, 243, \ldots$ is a geometric sequence. The numerator of the nth term is 3^n The sequence of denominators $4, 7, 10, 13, 16, \ldots$ is an arithmetic sequence. The denominator of the nth term is $4 + 3(n-1) = 3n+1$. Therefore, we can describe the sequence by the explicit formula $a_n = \frac{3^n}{3n+1.}$

Exercise 5.1.1

Find an explicit formula for the nth term of the sequence $\left\{\frac{1}{5}, -\frac{1}{7}, \frac{1}{9}, -\frac{1}{11}, \ldots\right\}$.

Hint

The denominators form an arithmetic sequence.

Answer

$$a_n = \frac{(-1)^{n+1}}{3+2n}$$

Example 5.1.2: Defined by Recurrence Relations

For each of the following recursively defined sequences, find an explicit formula for the sequence.

a. $a_1 = 2,\ a_n = -3a_{n-1}$ for $n \geq 2$

b. $a_1 = \left(\frac{1}{2}\right),\ a_n = a_{n-1} + \left(\frac{1}{2}\right)^n$ for $n \geq 2$

Solution

a. Writing out the first few terms, we have

$$\begin{aligned} a_1 &= 2 \\ a_2 &= -3a_1 = -3(2) \\ a_3 &= -3a_2 = (-3)^2 2 \\ a_4 &= -3a_3 = (-3)^3 2. \end{aligned}$$

In general,

$$a_n = 2(-3)^{n-1}.$$

b. Write out the first few terms:

$$\begin{aligned} a_1 &= \frac{1}{2} \\ a_2 &= a_1 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \\ a_3 &= a_2 + \left(\frac{1}{2}\right)^3 = \frac{3}{4} + \frac{1}{8} = \frac{7}{8} \\ a_4 &= a_3 + \left(\frac{1}{2}\right)^4 = \frac{7}{8} + \frac{1}{16} = \frac{15}{16} \quad . \end{aligned}$$

From this pattern, we derive the explicit formula

$$a_n = \frac{2^n - 1}{2^n} = 1 - \frac{1}{2^n}\ .$$

Exercise 5.1.2

Find an explicit formula for the sequence defined recursively such that $a_1 = -4$ and $a_n = a_{n-1} + 6$.

Hint

This is an arithmetic sequence.

Answer

$a_n = 6n - 10$

Limit of a Sequence

A fundamental question that arises regarding infinite sequences is the behavior of the terms as n gets larger. Since a sequence is a function defined on the positive integers, it makes sense to discuss the limit of the terms as $n \to \infty$. For example, consider the following four sequences and their different behaviors as $n \to \infty$ (Figure 5.1.2):

a. $\{1 + 3n\} = \{4, 7, 10, 13, \ldots\}$.The terms $1 + 3n$ become arbitrarily large as $n \to \infty$. In this case, we say that $1 + 3n \to \infty$ as $n \to \infty$.

b. $\left\{1 - \left(\frac{1}{2}\right)^n\right\} = \left\{\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16} \ldots\right\}$. The terms $1 - \left(\frac{1}{2}\right)^n \to 1$ as $n \to \infty$.

c. $\{(-1)^n\} = \{-1, 1, -1, 1, \ldots\}$.The terms alternate but do not approach one single value as $n \to \infty$.

d. $\left\{\dfrac{(-1)^n}{n}\right\} = \left\{-1, \dfrac{1}{2}, -\dfrac{1}{3}, \dfrac{1}{4}, \ldots\right\}$. The terms alternate for this sequence as well, but $\dfrac{(-1)^n}{n} \to 0$ as $n \to \infty$.

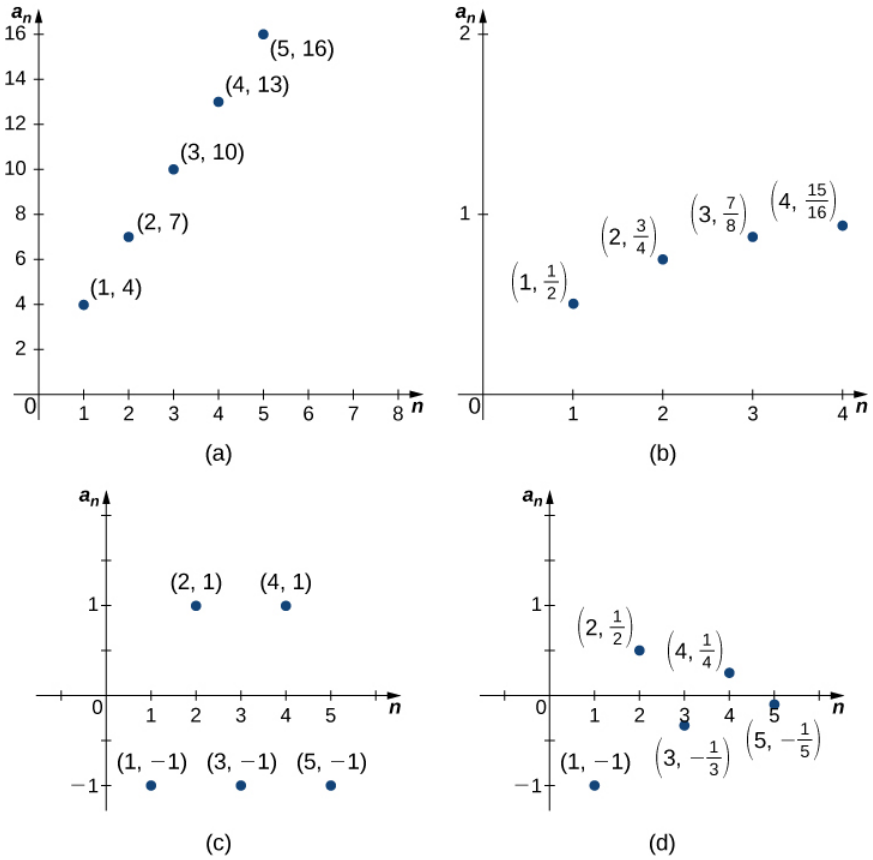


Figure 5.1.2: (a) The terms in the sequence become arbitrarily large as $n \to \infty$. (b) The terms in the sequence approach 1 as $n \to \infty$. (c) The terms in the sequence alternate between 1 and -1 as $n \to \infty$. (d) The terms in the sequence alternate between positive and negative values but approach 0 as $n \to \infty$.

From these examples, we see several possibilities for the behavior of the terms of a sequence as $n \to \infty$. In two of the sequences, the terms approach a finite number as $n \to \infty$. In the other two sequences, the terms do not. If the terms of a sequence approach a finite number L as $n \to \infty$, we say that the sequence is a convergent sequence and the real number L is the limit of the sequence. We can give an informal definition here.

Definition: Convergent and Divergent Sequences

Given a sequence $\{a_n\}$, if the terms a_n become arbitrarily close to a finite number L as n becomes sufficiently large, we say $\{a_n\}$ is a **convergent sequence** and L is the **limit of the sequence.** In this case, we write

$$\lim_{n\to\infty} a_n = L.$$

If a sequence $\{a_n\}$ is not convergent, we say it is a **divergent sequence.**

From Figure $5.1.2b$, we see that the terms in the sequence $\left\{1 - \left(\dfrac{1}{2}\right)^n\right\}$ are becoming arbitrarily close to 1 as n becomes very large. We conclude that $\left\{1 - \left(\dfrac{1}{2}\right)^n\right\}$ is a convergent sequence and its limit is 1. In contrast, from Figure $5.1.2a$, we see that the terms in the sequence $\{1 + 3n\}$ are not approaching a finite number as n becomes larger. We say that $\{1 + 3n\}$ is a divergent sequence.

In the informal definition for the limit of a sequence, we used the terms “arbitrarily close” and “sufficiently large.” Although these phrases help illustrate the meaning of a converging sequence, they are somewhat vague. To be more precise, we now present the

more formal definition of limit for a sequence and show these ideas graphically in Figure 5.1.3.

Definition: Convergence

A sequence $\{a_n\}$ converges to a real number L if for all $\varepsilon > 0$, there exists an integer N such that for all $n \geq N$ $|a_n - L| < \varepsilon$. The number L is the limit of the sequence and we write

$$\lim_{n\to\infty} a_n = L \text{ or } a_n \to L.$$

In this case, we say the sequence $\{a_n\}$ is a convergent sequence. If a sequence does not converge, it is a divergent sequence, and we say the limit does not exist.

We remark that the convergence or divergence of a sequence $\{a_n\}$ depends only on what happens to the terms a_n as $n \to \infty$. Therefore, if a finite number of terms $b_1, b_2, \ldots, b_N$ are placed before a_1 to create a new sequence

$$b_1,\ b_2,\ \ldots,\ b_N,\ a_1,\ a_2,\ \ldots,$$

this new sequence will converge if $\{a_n\}$ converges and diverge if $\{a_n\}$ diverges. Further, if the sequence $\{a_n\}$ converges to L, this new sequence will also converge to L.

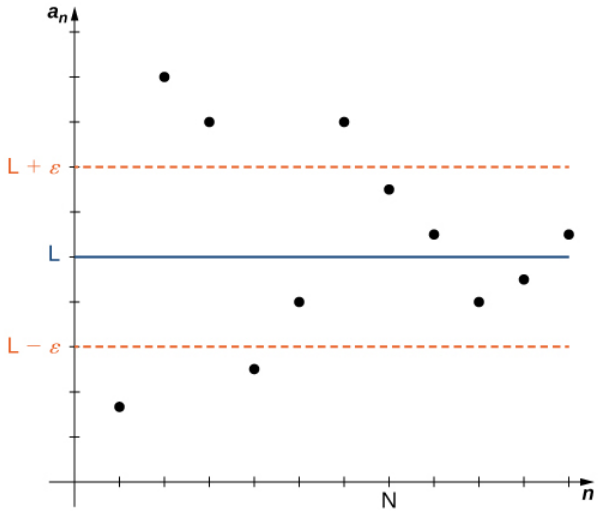


Figure 5.1.3: As n increases, the terms a_n become closer to L. For values of $n \geq N$, the distance between each point (n, a_n) and the line $y = L$ is less than ε.

As defined above, if a sequence does not converge, it is said to be a divergent sequence. For example, the sequences $\{1+3n\}$ and $\{(-1)^n\}$ shown in Figure 5.1.2 diverge. However, different sequences can diverge in different ways. The sequence $\{(-1)^n\}$ diverges because the terms alternate between 1 and -1, but do not approach one value as $n \to \infty$. On the other hand, the sequence $\{1+3n\}$ diverges because the terms $1+3n \to \infty$ as $n \to \infty$. We say the sequence $\{1+3n\}$ diverges to infinity and write $\lim_{n\to\infty}(1+3n) = \infty$. It is important to recognize that this notation does not imply the limit of the sequence $\{1+3n\}$ exists. The sequence is, in fact, divergent. Writing that the limit is infinity is intended only to provide more information about why the sequence is divergent. A sequence can also diverge to negative infinity. For example, the sequence $\{-5n+2\}$ diverges to negative infinity because $-5n+2 \to -\infty$ as $n \to -\infty$. We write this as $\lim_{n\to\infty}(-5n+2) =\to -\infty$.

Because a sequence is a function whose domain is the set of positive integers, we can use properties of limits of functions to determine whether a sequence converges. For example, consider a sequence $\{a_n\}$ and a related function f defined on all positive real numbers such that $f(n) = a_n$ for all integers $n \geq 1$. Since the domain of the sequence is a subset of the domain of f, if $\lim_{x\to\infty} f(x)$ exists, then the sequence converges and has the same limit. For example, consider the sequence $\left\{\frac{1}{n}\right\}$ and the related function $f(x) = \frac{1}{x}$. Since the function f defined on all real numbers $x > 0$ satisfies $f(x) = \frac{1}{x} \to 0$ as $x \to \infty$, the sequence $\left\{\frac{1}{n}\right\}$ must satisfy $\frac{1}{n} \to 0$ as $n \to \infty$.

Limit of a Sequence Defined by a Function

Consider a sequence $\{a_n\}$ such that $a_n = f(n)$ for all $n \geq 1$. If there exists a real number L such that

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

then $\{a_n\}$ converges and

$$\lim_{n\to\infty} a_n = L.$$

We can use this theorem to evaluate $\lim_{n\to\infty} r^n$ for $0 \leq r \leq 1$. For example, consider the sequence $\{(1/2)^n\}$ and the related exponential function $f(x) = (1/2)^x$. Since $\lim_{x\to\infty} (1/2)^x = 0$, we conclude that the sequence $\{(1/2)^n\}$ converges and its limit is 0. Similarly, for any real number r such that $0 \leq r < 1$, $\lim_{x\to\infty} r^x = 0$, and therefore the sequence $\{r^n\}$ converges. On the other hand, if $r = 1$, then $\lim_{x\to\infty} r^x = 1$, and therefore the limit of the sequence $\{1^n\}$ is 1. If $r > 1$, $\lim_{x\to\infty} r^x = \infty$, and therefore we cannot apply this theorem. However, in this case, just as the function r^x grows without bound as $n \to \infty$, the terms r^n in the sequence become arbitrarily large as $n \to \infty$, and we conclude that the sequence $\{r^n\}$ diverges to infinity if $r > 1$.

We summarize these results regarding the geometric sequence r^n:

$$r^n \to 0 \text{ if } 0 < r < 1$$

$$r^n \to 1 \text{ if } r = 1$$

$$r^n \to \infty \text{ if } r > 1.$$

Later in this section we consider the case when $r < 0$.

We now consider slightly more complicated sequences. For example, consider the sequence $\{(2/3)^n + (1/4)^n\}$. The terms in this sequence are more complicated than other sequences we have discussed, but luckily the limit of this sequence is determined by the limits of the two sequences $\{(2/3)^n\}$ and $\{(1/4)^n\}$. As we describe in the following algebraic limit laws, since $\{(2/3)^n\}$ and $\{1/4)^n\}$ both converge to 0, the sequence $\{(2/3)^n + (1/4)^n\}$ converges to $0 + 0 = 0$. Just as we were able to evaluate a limit involving an algebraic combination of functions f and g by looking at the limits of f and g (see Introduction to Limits), we are able to evaluate the limit of a sequence whose terms are algebraic combinations of a_n and b_n by evaluating the limits of $\{a_n\}$ and $\{b_n\}$.

Algebraic Limit Laws

Given sequences $\{a_n\}$ and $\{b_n\}$ and any real number c, if there exist constants A and B such that $\lim_{n\to\infty} a_n = A$ and $\lim_{n\to\infty} b_n = B$, then

i. $\lim_{n\to\infty} c = c$

ii. $\lim_{n\to\infty} ca_n = c \lim_{n\to\infty} a_n = cA$

iii. $\lim_{n\to\infty} (a_n \pm b_n) = \lim_{n\to\infty} a_n \pm \lim_{n\to\infty} b_n = A \pm B$

iv. $\lim_{n\to\infty} (a_n \cdot b_n) = \left(\lim_{n\to\infty} a_n\right) \cdot \left(\lim_{n\to\infty} b_n\right) = A \cdot B$

v. $\lim_{n\to\infty} \left(\frac{a_n}{b_n}\right) = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n} = \frac{A}{B}$, provided $B \neq 0$ and each $b_n \neq 0$.

Proof

We prove part iii.

Let $\epsilon > 0$. Since $\lim_{n\to\infty} a_n = A$, there exists a constant positive integer N_1 such that for all $n \geq N_1$. Since $\lim_{n\to\infty} b_n = B$, there exists a constant N_2 such that $|b_n - B| < \varepsilon/2$ for all $n \geq N_2$. Let N be the largest of N_1 and N_2. Therefore, for all $n \geq N$, $|(a_n + b_n) - (A + B)| \leq |a_n - A| + |b_n - B| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$.

□

The algebraic limit laws allow us to evaluate limits for many sequences. For example, consider the sequence $a_n = \frac{1}{n^2}$. As shown earlier, $\lim_{n\to\infty} \frac{1}{n} = 0$. Similarly, for any positive integer k, we can conclude that

$$\lim_{n\to\infty} \frac{1}{n^k} = 0.$$

In the next example, we make use of this fact along with the limit laws to evaluate limits for other sequences.

Example 5.1.3: Determining Convergence and Finding Limits

For each of the following sequences, determine whether or not the sequence converges. If it converges, find its limit.

a. $\left\{5 - \frac{3}{n^2}\right\}$

b. $\left\{\frac{3n^4 - 7n^2 + 5}{6 - 4n^4}\right\}$

c. $\left\{\frac{2^n}{n^2}\right\}$

d. $\left\{\left(1 + \frac{4}{n}\right)^n\right\}$

Solution

a. We know that $\lim_{n\to\infty} \frac{1}{n} = 0$. Using this fact, we conclude that

$$\lim_{n\to\infty} \frac{1}{n^2} = \lim_{n\to\infty} \frac{1}{n} . \lim_{n\to\infty} \frac{1}{n} = 0.$$

Therefore,

$$\lim_{n\to\infty} \left(5 - \frac{3}{n^2}\right) = \lim_{n\to\infty} 5 - 3 \lim_{n\to\infty} \frac{1}{n^2} = 5 - 3.0 = 5.$$

The sequence converges and its limit is 5.

b. By factoring n^4 out of the numerator and denominator and using the limit laws above, we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{3n^4 - 7n^2 + 5}{6 - 4n^4} &= \lim_{n\to\infty} \frac{3 - \frac{7}{n^2} + \frac{5}{n^4}}{\frac{6}{n^4} - 4} \\
&= \frac{\lim_{n\to\infty} (3 - \frac{7}{n^2} + \frac{5}{n^4})}{\lim_{n\to\infty} (\frac{6}{n^4} - 4)} \\
&= \frac{\lim_{n\to\infty} (3) - \lim_{n\to\infty} \frac{7}{n^2} + \lim_{n\to\infty} \frac{5}{n^4}}{\lim_{n\to\infty} \frac{6}{n^4} - \lim_{n\to\infty} (4)} \\
&= \frac{\lim_{n\to\infty} (3) - 7 \cdot \lim_{n\to\infty} \frac{1}{n^2} + 5 \cdot \lim_{n\to\infty} \frac{1}{n^4}}{6 \cdot \lim_{n\to\infty} \frac{1}{n^4} - \lim_{n\to\infty} (4)} \\
&= \frac{3 - 7 \cdot 0 + 5 \cdot 0}{6 \cdot 0 - 4} = -\frac{3}{4}.
\end{aligned}$$

The sequence converges and its limit is $-3/4$.

c. Consider the related function $f(x) = 2^x/x^2$ defined on all real numbers $x > 0$. Since $2^x \to \infty$ and $x^2 \to \infty$ as $x \to \infty$, apply L'Hôpital's rule and write

$$\begin{aligned}\lim_{x\to\infty} \frac{2^x}{x^2} &= \lim_{x\to\infty} \frac{2^x \ln 2}{2x} && \text{Take the derivatives of the numerator and denominator.}\\ &= \lim_{x\to\infty} \frac{2^x (\ln 2)^2}{2} && \text{Take the derivatives again.}\\ &= \infty.\end{aligned}$$

We conclude that the sequence diverges.

d. Consider the function $f(x) = \left(1 + \frac{4}{x}\right)^x$ defined on all real numbers $x > 0$. This function has the indeterminate form 1^∞ as $x \to \infty$. Let

$$y = \lim_{x\to\infty} \left(1 + \frac{4}{x}\right)^x.$$

Now taking the natural logarithm of both sides of the equation, we obtain

$$\ln(y) = \ln\left[\lim_{x\to\infty} \left(1 + \frac{4}{x}\right)^x\right].$$

Since the function $f(x) = \ln(x)$ is continuous on its domain, we can interchange the limit and the natural logarithm. Therefore,

$$\ln(y) = \lim_{x\to\infty} \left[\ln\left(1 + \frac{4}{x}\right)^x\right].$$

Using properties of logarithms, we write

$$\lim_{x\to\infty} \left[\ln\left(1 + \frac{4}{x}\right)^x\right] = \lim_{x\to\infty} x \ln\left(1 + \frac{4}{x}\right).$$

Since the right-hand side of this equation has the indeterminate form $\infty \cdot 0$, rewrite it as a fraction to apply L'Hôpital's rule. Write

$$\lim_{x\to\infty} x \ln\left(1 + \frac{4}{x}\right) = \lim_{x\to\infty} \frac{\ln(1 + 4/x)}{1/x}.$$

Since the right-hand side is now in the indeterminate form 0/0, we are able to apply L'Hôpital's rule. We conclude that

$$\lim_{x\to\infty} \frac{\ln(1 + 4/x)}{1/x} = \lim_{x\to\infty} \frac{4}{1 + 4/x} = 4.$$

Therefore, $\ln(y) = 4$ and $y = e^4$. Therefore, since $\lim_{x\to\infty} \left(1 + \frac{4}{x}\right)^x = e^4$, we can conclude that the sequence $\left\{\left(1 + \frac{4}{n}\right)^n\right\}$ converges to e^4.

? Exercise 5.1.3

Consider the sequence $\left\{(5n^2 + 1)/e^n\right\}$. Determine whether or not the sequence converges. If it converges, find its limit.

Hint

Use L'Hôpital's rule.

Answer

The sequence converges, and its limit is 0

Recall that if f is a continuous function at a value L, then $f(x) \to f(L)$ as $x \to L$. This idea applies to sequences as well. Suppose a sequence $a_n \to L$, and a function f is continuous at L. Then $f(a_n) \to f(L)$. This property often enables us to find limits for complicated sequences. For example, consider the sequence $\sqrt{5 - \frac{3}{n^2}}$. From Example a. we know the sequence $5 - \frac{3}{n^2} \to 5$. Since $\sqrt{x}$ is a continuous function at $x = 5$,

$$\lim_{n\to\infty} \sqrt{5 - \frac{3}{n^2}} = \sqrt{\lim_{n\to\infty} (5 - \frac{3}{n^2})} = \sqrt{5}.$$

Continuous Functions Defined on Convergent Sequences

Consider a sequence $\{a_n\}$ and suppose there exists a real number L such that the sequence $\{a_n\}$ converges to L. Suppose f is a continuous function at L. Then there exists an integer N such that f is defined at all values an for $n \geq N$, and the sequence $\{f(a_n)\}$ converges to $f(L)$ (Figure 5.1.4).

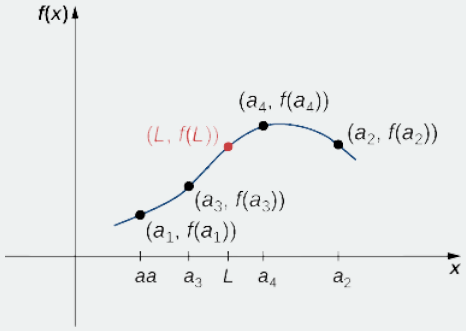


Figure 5.1.4: Because f is a continuous function as the inputs $a_1, a_2, a_3, \ldots$ approach L, the outputs $f(a_1), f(a_2), f(a_3), \ldots$ approach $f(L)$.

Proof

Let $\epsilon > 0$. Since f is continuous at L, there exists $\delta > 0$ such that $|f(x) - f(L)| < \varepsilon$ if $|x - L| < \delta$. Since the sequence $\{a_n\}$ converges to L, there exists N such that $|a_n - L| < \delta$ for all $n \geq N$. Therefore, for all $n \geq N$, $|a_n - L| < \delta$, which implies $|f(a_n) - f(L)| < \varepsilon$. We conclude that the sequence $\{f(a_n)\}$ converges to $f(L)$.

□

Example 5.1.4: Limits Involving Continuous Functions Defined on Convergent Sequences

Determine whether the sequence $\{\cos(3/n^2)\}$ converges. If it converges, find its limit.

Solution:

Since the sequence $\{3/n^2\}$ converges to 0 and $\cos x$ is continuous at $x = 0$, we can conclude that the sequence $\{\cos(3/n^2)\}$ converges and

$$\lim_{n\to\infty} \cos\left(\frac{3}{n^2}\right) = \cos 0 = 1.$$

Exercise 5.1.4

Determine if the sequence $\left\{\sqrt{\frac{2n+1}{3n+5}}\right\}$ converges. If it converges, find its limit.

Hint

Consider the sequence $\left\{\frac{2n+1}{3n+5}\right\}$.

Answer

The sequence converges, and its limit is $\sqrt{2/3}$.

Another theorem involving limits of sequences is an extension of the Squeeze Theorem for limits discussed in Introduction to Limits.

Squeeze Theorem for Sequences

Consider sequences $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$. Suppose there exists an integer N such that

$$a_n \leq b_n \leq c_n \text{ for all } n \geq N.$$

If there exists a real number L such that

$$\lim_{n\to\infty} a_n = L = \lim_{n\to\infty} c_n,$$

then $\{b_n\}$ converges and $\lim_{n\to\infty} b_n = L$ (Figure 5.1.5).

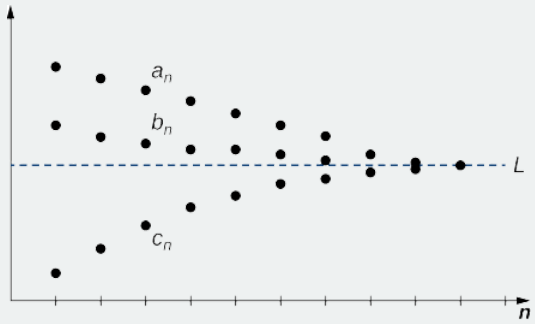


Figure 5.1.5: Each term bn satisfies $a_n \leq b_n \leq c_n$ and the sequences $\{a_n\}$ and $\{c_n\}$ converge to the same limit, so the sequence $\{b_n\}$ must converge to the same limit as well.

Proof

Let $\varepsilon > 0$. Since the sequence $\{a_n\}$ converges to L, there exists an integer N_1 such that $|a_n - L| < \varepsilon$ for all $n \geq N_1$. Similarly, since $\{c_n\}$ converges to L, there exists an integer N_2 such that $|c_n - L| < \varepsilon$ for all $n \geq N_2$. By assumption, there exists an integer N such that $a_n \leq b_n \leq c_n$ for all $n \geq N$. Let M be the largest of N_1, N_2, and N. We must show that $|b_n - L| < \varepsilon$ for all $n \geq M$. For all $n \geq M$,

$$-\varepsilon < -|a_n - L| \leq a_n - L \leq b_n - L \leq c_n - L \leq |c_n - L| < \varepsilon$$

Therefore, $-\varepsilon < b_n - L < \varepsilon$, and we conclude that $|b_n - L| < \varepsilon$ for all $n \geq M$, and we conclude that the sequence b_n converges to L.

□

Example 5.1.5: Using the Squeeze Theorem

Use the Squeeze Theorem to find the limit of each of the following sequences.

a. $\left\{\dfrac{\cos n}{n^2}\right\}$

b. $\left\{\left(-\dfrac{1}{2}\right)^n\right\}$

Solution

a. Since $-1 \leq \cos n \leq 1$ for all integers n, we have

$$-\frac{1}{n^2} \leq \frac{\cos n}{n^2} \leq \frac{1}{n^2}.$$

Since $-1/n^2 \to 0$ and $1/n^2 \to 0$, we conclude that $\cos n/n^2 \to 0$ as well.

b. Since

$$-\frac{1}{2^n} \leq \left(-\frac{1}{2}\right)^n \leq \frac{1}{2^n}$$

for all positive integers n, $-1/2^n \to 0$ and $1/2^n \to 0$, we can conclude that $(-1/2)^n \to 0$.

Exercise 5.1.5

Find $\lim\limits_{n\to\infty} \dfrac{2n-\sin n}{n}$.

Hint

Use the fact that $-1 \leq \sin n \leq 1$.

Answer

2

Using the idea from Example 5.1.5 b we conclude that $r^n \to 0$ for any real number r such that $-1 < r < 0$. If $r < -1$, the sequence r^n diverges because the terms oscillate and become arbitrarily large in magnitude. If $r = -1$, the sequence $r^n = (-1)^n$ diverges, as discussed earlier. Here is a summary of the properties for geometric sequences.

$$r^n \to 0 \text{ if } |r| < 1$$

$$r^n \to 1 \text{ if } r = 1$$

$$r^n \to \infty \text{ if } r > 1$$

$$\{r^n\} \text{ diverges if } r \leq -1$$

Bounded Sequences

We now turn our attention to one of the most important theorems involving sequences: the Monotone Convergence Theorem. Before stating the theorem, we need to introduce some terminology and motivation. We begin by defining what it means for a sequence to be bounded.

Definition: Bound Sequences

A sequence $\{a_n\}$ is **bounded above** if there exists a real number M such that

$$a_n \leq M$$

for all positive integers n.

A sequence $\{a_n\}$ is **bounded below** if there exists a real number m such that

$$m \leq a_n$$

for all positive integers n.

A sequence $\{a_n\}$ is a **bounded sequence** if it is bounded above and bounded below.

If a sequence is not bounded, it is an **unbounded sequence.**

For example, the sequence $\{1/n\}$ is bounded above because $1/n \leq 1$ for all positive integers n. It is also bounded below because $1/n \geq 0$ for all positive integers n. Therefore, $\{1/n\}$ is a bounded sequence. On the other hand, consider the sequence $\{2^n\}$. Because $2^n \geq 2$ for all $n \geq 1$, the sequence is bounded below. However, the sequence is not bounded above. Therefore, $\{2^n\}$ is an unbounded sequence.

We now discuss the relationship between boundedness and convergence. Suppose a sequence $\{a_n\}$ is unbounded. Then it is not bounded above, or not bounded below, or both. In either case, there are terms an that are arbitrarily large in magnitude as n gets

larger. As a result, the sequence $\{a_n\}$ cannot converge. Therefore, being bounded is a necessary condition for a sequence to converge.

Convergent Sequences Are Bounded

If a sequence $\{a_n\}$ converges, then it is bounded.

Note that a sequence being bounded is not a sufficient condition for a sequence to converge. For example, the sequence $\{(-1)^n\}$ is bounded, but the sequence diverges because the sequence oscillates between 1 and -1 and never approaches a finite number. We now discuss a sufficient (but not necessary) condition for a bounded sequence to converge.

Consider a bounded sequence $\{a_n\}$. Suppose the sequence $\{a_n\}$ is increasing. That is, $a_1 \leq a_2 \leq a_3 \ldots$. Since the sequence is increasing, the terms are not oscillating. Therefore, there are two possibilities. The sequence could diverge to infinity, or it could converge. However, since the sequence is bounded, it is bounded above and the sequence cannot diverge to infinity. We conclude that $\{a_n\}$ converges. For example, consider the sequence

$$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots\right\}.$$

Since this sequence is increasing and bounded above, it converges. Next, consider the sequence

$$\left\{2, 0, 3, 0, 4, 0, 1, -\frac{1}{2}, -\frac{1}{3}, -\frac{1}{4}, \ldots\right\}.$$

Even though the sequence is not increasing for all values of n, we see that $-1/2 < -1/3 < -1/4 < \cdots$. Therefore, starting with the eighth term, $a_8 = -1/2$, the sequence is increasing. In this case, we say the sequence is eventually increasing. Since the sequence is bounded above, it converges. It is also true that if a sequence is decreasing (or eventually decreasing) and bounded below, it also converges.

Definition

A sequence $\{a_n\}$ is increasing for all $n \geq n_0$ if

$$a_n \leq a_{n+1} \text{ for all } n \geq n_0.$$

A sequence $\{a_n\}$ is decreasing for all $n \geq n_0$ if

$$a_n \geq a_{n+1} \text{ for all } n \geq n_0.$$

A sequence $\{a_n\}$ is a *monotone sequence* for all $n \geq n_0$ if it is increasing for all $n \geq n_0$ or decreasing for all $n \geq n_0$.

We now have the necessary definitions to state the Monotone Convergence Theorem, which gives a sufficient condition for convergence of a sequence.

Definition: Monotone Convergence Theorem

If $\{a_n\}$ is a bounded sequence and there exists a positive integer n_0 such that $\{a_n\}$ is monotone for all $n \geq n_0$, then $\{a_n\}$ converges.

The proof of this theorem is beyond the scope of this text. Instead, we provide a graph to show intuitively why this theorem makes sense (Figure 5.1.6).

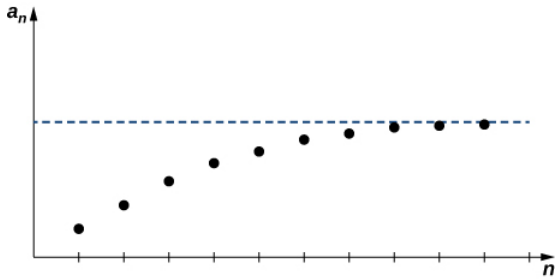


Figure 5.1.6: Since the sequence $\{a_n\}$ is increasing and bounded above, it must converge.

In the following example, we show how the Monotone Convergence Theorem can be used to prove convergence of a sequence.

Example 5.1.6: Using the Monotone Convergence Theorem

For each of the following sequences, use the Monotone Convergence Theorem to show the sequence converges and find its limit.

a. $\left\{\dfrac{4^n}{n!}\right\}$

b. $\{a_n\}$ defined recursively such that

$$a_1 = 2 \text{ and } a_{n+1} = \frac{a_n}{2} + \frac{1}{2a_n} \text{ for all } n \geq 2.$$

Solution

a. Writing out the first few terms, we see that

$$\left\{\frac{4^n}{n!}\right\} = \left\{4,\ 8,\ \frac{32}{3},\ \frac{32}{3},\ \frac{128}{15},\ \ldots\right\}.$$

At first, the terms increase. However, after the third term, the terms decrease. In fact, the terms decrease for all $n \geq 3$. We can show this as follows.

$$a_{n+1} = \frac{4^{n+1}}{(n+1)!} = \frac{4}{n+1} \cdot \frac{4^n}{n!} = \frac{4}{n+1} \cdot a_n \leq a_n \quad \text{if } n \geq 3.$$

Therefore, the sequence is decreasing for all $n \geq 3$. Further, the sequence is bounded below by 0 because $4n/n! \geq 0$ for all positive integers n. Therefore, by the Monotone Convergence Theorem, the sequence converges.

To find the limit, we use the fact that the sequence converges and let $L = \lim\limits_{n\to\infty} a_n$. Now note this important observation. Consider $\lim\limits_{n\to\infty} a_{n+1}$. Since

$$\{a_{n+1}\} = \{a_2,\ a_3,\ a_4,\ \ldots\},$$

the only difference between the sequences $\{a_{n+1}\}$ and $\{a_n\}$ is that $\{a_{n+1}\}$ omits the first term. Since a finite number of terms does not affect the convergence of a sequence,

$$\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} a_n = L.$$

Combining this fact with the equation

$$a_{n+1} = \frac{4}{n+1} a_n$$

and taking the limit of both sides of the equation

$$\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} \frac{4}{n+1} a_n,$$

we can conclude that

$$L = 0 \cdot L = 0.$$

b. Writing out the first several terms,

$$\left\{2, \frac{5}{4}, \frac{41}{40}, \frac{3281}{3280}, \ldots\right\}.$$

we can conjecture that the sequence is decreasing and bounded below by 1. To show that the sequence is bounded below by 1, we can show that

$$\frac{a_n}{2} + \frac{1}{2a_n} \geq 1.$$

To show this, first rewrite

$$\frac{a_n}{2} + \frac{1}{2a_n} = \frac{a_n^2 + 1}{2a_n} \ .$$

Since $a_1 > 0$ and a_2 is defined as a sum of positive terms, $a_2 > 0$. Similarly, all terms $a_n > 0$. Therefore,

$$\frac{a_n^2 + 1}{2a_n} \geq 1$$

if and only if

$$a_n^2 + 1 \geq 2a_n \ .$$

Rewriting the inequality $a_n^2 + 1 \geq 2a_n$ as $a_n^2 - 2a_n + 1 \geq 0$, and using the fact that

$$a_n^2 - 2a_n + 1 = (a_n - 1)^2 \geq 0$$

because the square of any real number is nonnegative, we can conclude that

$$\frac{a^n}{2} + \frac{1}{2a_n} \geq 1.$$

To show that the sequence is decreasing, we must show that $a_{n+1} \leq a_n$ for all $n \geq 1$. Since $1 \leq a_n^2$, it follows that

$$a_n^2 + 1 \leq 2a_n^2 \ .$$

Dividing both sides by $2a_n$, we obtain

$$\frac{a_n}{2} + \frac{1}{2a_n} \leq a_n.$$

Using the definition of a_{n+1}, we conclude that

$$a_{n+1} = \frac{a_n}{2} + \frac{1}{2a_n} \leq a_n \ .$$

Since $\{a_n\}$ is bounded below and decreasing, by the Monotone Convergence Theorem, it converges.

To find the limit, let $L = \lim_{n\to\infty} a_n$. Then using the recurrence relation and the fact that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n+1}$, we have

$$\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} \left(\frac{a_n}{2} + \frac{1}{2a_n}\right),$$

and therefore

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

Multiplying both sides of this equation by $2L$, we arrive at the equation

$$2L^2 = L^2 + 1 \, .$$

Solving this equation for L, we conclude that $L^2 = 1$, which implies $L = \pm 1$. Since all the terms are positive, the limit $L = 1$.

Exercise 5.1.6

Consider the sequence $\{a_n\}$ defined recursively such that $a_1 = 1$, $a_n = \dfrac{a_{n-1}}{2}$. Use the Monotone Convergence Theorem to show that this sequence converges and find its limit.

Hint

Show the sequence is decreasing and bounded below.

Answer

0.

Definition: Fibonacci Numbers

The *Fibonacci numbers* are defined recursively by the sequence $\{F_n\}$ where $F_0 = 0$, $F_1 = 1$ and for $n \geq 2$,

$$F_n = F_{n-1} + F_{n-2}.$$

Here we look at properties of the Fibonacci numbers.

1. Write out the first twenty Fibonacci numbers.

2. Find a closed formula for the Fibonacci sequence by using the following steps.

   a. Consider the recursively defined sequence x_n where $x_0 = c$ and $x_{n+1} = ax_n$. Show that this sequence can be described by the closed formula $x_n = ca^n$ for all $n \geq 0$.

   b. Using the result from part a. as motivation, look for a solution of the equation

   $F_n = F_{n-1} + F_{n-2}$

   of the form $F_n = c\lambda^n$. Determine what two values for λ will allow F_n to satisfy this equation.

   c. Consider the two solutions from part b.: λ_1 and λ_2. Let $F_n = c_1\lambda_1^n + c_2\lambda_2^n$. Use the initial conditions F_0 and F_1 to determine the values for the constants c_1 and c_2 and write the closed formula F_n.

3. Use the answer in 2 c. to show that

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \frac{1+\sqrt{5}}{2}.$$

The number $\phi = (1+\sqrt{5})/2$ is known as the **golden ratio** (Figure 5.1.7 and Figure 5.1.8).



Figure 5.1.7: The seeds in a sunflower exhibit spiral patterns curving to the left and to the right. The number of spirals in each direction is always a Fibonacci number—always. (credit: modification of work by Esdras Calderan, Wikimedia Commons)

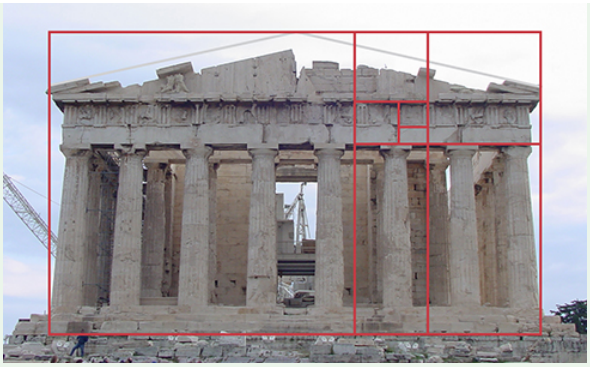


Figure 5.1.8: The proportion of the golden ratio appears in many famous examples of art and architecture. The ancient Greek temple known as the Parthenon was designed with these proportions, and the ratio appears again in many of the smaller details. (credit: modification of work by TravelingOtter, Flickr).

Key Concepts

- To determine the convergence of a sequence given by an explicit formula $a_n = f(n)$, we use the properties of limits for functions.
- If $\{a_n\}$ and $\{b_n\}$ are convergent sequences that converge to A and B, respectively, and c is any real number, then the sequence $\{ca_n\}$converges to $c \cdot A$, the sequences $\{a_n \pm b_n\}$ converge to $A \pm B$, the sequence $\{a_n \cdot b_n\}$ converges to $A \cdot B$, and the sequence $\{a_n / b_n\}$ converges to A/B, provided $B \neq 0$.
- If a sequence is bounded and monotone, then it converges, but not all convergent sequences are monotone.
- If a sequence is unbounded, it diverges, but not all divergent sequences are unbounded.
- The geometric sequence $\{r^n\}$ converges if and only if $|r| < 1$ or $r = 1$.

Glossary

arithmetic sequence

a sequence in which the difference between every pair of consecutive terms is the same is called an arithmetic sequence

bounded above

a sequence $\{a_n\}$ is bounded above if there exists a constant M such that $a_n \leq M$ for all positive integers n

bounded below

a sequence $\{a_n\}$ is bounded below if there exists a constant M such that $M \leq a_n$ for all positive integers n

bounded sequence

a sequence $\{a_n\}$ is bounded if there exists a constant M such that $|a_n| \leq M$ for all positive integers n

convergent sequence

a convergent sequence is a sequence $\{a_n\}$ for which there exists a real number L such that a_n is arbitrarily close to L as long as n is sufficiently large

divergent sequence

a sequence that is not convergent is divergent

explicit formula

a sequence may be defined by an explicit formula such that $a_n = f(n)$

geometric sequence

a sequence $\{a_n\}$ in which the ratio a_{n+1}/a_n is the same for all positive integers n is called a geometric sequence

index variable

the subscript used to define the terms in a sequence is called the index

limit of a sequence

the real number L to which a sequence converges is called the limit of the sequence

monotone sequence

an increasing or decreasing sequence

recurrence relation

a recurrence relation is a relationship in which a term a_n in a sequence is defined in terms of earlier terms in the sequence

sequence

an ordered list of numbers of the form $a_1, a_2, a_3, \ldots$ is a sequence

term

the number a_n in the sequence $\{a_n\}$ is called the nth term of the sequence

unbounded sequence

a sequence that is not bounded is called unbounded

This page titled 5.1: Sequences is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.1: Sequences** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

5.2: Infinite Series

Learning Objectives

- Explain the meaning of the sum of an infinite series.
- Calculate the sum of a geometric series.
- Evaluate a telescoping series.

We have seen that a sequence is an ordered set of terms. If you add these terms together, you get a series. In this section we define an infinite series and show how series are related to sequences. We also define what it means for a series to converge or diverge. We introduce one of the most important types of series: the geometric series. We will use geometric series in the next chapter to write certain functions as polynomials with an infinite number of terms. This process is important because it allows us to evaluate, differentiate, and integrate complicated functions by using polynomials that are easier to handle. We also discuss the harmonic series, arguably the most interesting divergent series because it just fails to converge.

Sums and Series

An infinite series is a sum of infinitely many terms and is written in the form

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots.$$

But what does this mean? We cannot add an infinite number of terms in the same way we can add a finite number of terms. Instead, the value of an infinite series is defined in terms of the limit of partial sums. A partial sum of an infinite series is a finite sum of the form

$$\sum_{n=1}^{k} a_n = a_1 + a_2 + a_3 + \cdots + a_k.$$

To see how we use partial sums to evaluate infinite series, consider the following example. Suppose oil is seeping into a lake such that 1000 gallons enters the lake the first week. During the second week, an additional 500 gallons of oil enters the lake. The third week, 250 more gallons enters the lake. Assume this pattern continues such that each week half as much oil enters the lake as did the previous week. If this continues forever, what can we say about the amount of oil in the lake? Will the amount of oil continue to get arbitrarily large, or is it possible that it approaches some finite amount? To answer this question, we look at the amount of oil in the lake after k weeks. Letting S_k denote the amount of oil in the lake (measured in thousands of gallons) after k weeks, we see that

$$\begin{aligned}
S_1 &= 1 \\
S_2 &= 1 + 0.5 = 1 + \tfrac{1}{2} \\
S_3 &= 1 + 0.5 + 0.25 = 1 + \tfrac{1}{2} + \tfrac{1}{4} \\
S_4 &= 1 + 0.5 + 0.25 + 0.125 = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} \\
S_5 &= 1 + 0.5 + 0.25 + 0.125 + 0.0625 = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16}.
\end{aligned}$$

Looking at this pattern, we see that the amount of oil in the lake (in thousands of gallons) after k weeks is

$$S_k = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{2^{k-1}} = \sum_{n=1}^{k} \left(\frac{1}{2}\right)^{n-1}.$$

We are interested in what happens as $k \to \infty$. Symbolically, the amount of oil in the lake as $k \to \infty$ is given by the infinite series

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots.$$

At the same time, as $k \to \infty$, the amount of oil in the lake can be calculated by evaluating $\lim_{k\to\infty} S_k$. Therefore, the behavior of the infinite series can be determined by looking at the behavior of the sequence of partial sums S_k. If the sequence of partial sums S_k

converges, we say that the infinite series converges, and its sum is given by $\lim_{k\to\infty} S_k$. If the sequence S_k diverges, we say the infinite series diverges. We now turn our attention to determining the limit of this sequence S_k.

First, simplifying some of these partial sums, we see that

$$S_1 = 1$$
$$S_2 = 1 + \tfrac{1}{2} = \tfrac{3}{2}$$
$$S_3 = 1 + \tfrac{1}{2} + \tfrac{1}{4} = \tfrac{7}{4}$$
$$S_4 = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} = \tfrac{15}{8}$$
$$S_5 = 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16} = \tfrac{31}{16}.$$

Plotting some of these values in Figure 5.2.1, it appears that the sequence S_k could be approaching 2.

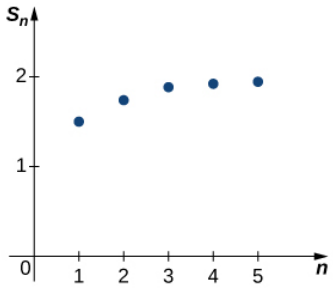


Figure 5.2.1: The graph shows the sequence of partial sums S_k. It appears that the sequence is approaching the value 2.

Let's look for more convincing evidence. In the following table, we list the values of S_k for several values of k.

k	5	10	15	20
S_k	1.9375	1.998	1.999939	1.999998

These data supply more evidence suggesting that the sequence S_k converges to 2. Later we will provide an analytic argument that can be used to prove that $\lim_{k\to\infty} S_k = 2$. For now, we rely on the numerical and graphical data to convince ourselves that the sequence of partial sums does actually converge to 2. Since this sequence of partial sums converges to 2, we say the infinite series converges to 2 and write

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} = 2.$$

Returning to the question about the oil in the lake, since this infinite series converges to 2, we conclude that the amount of oil in the lake will get arbitrarily close to 2000 gallons as the amount of time gets sufficiently large.

This series is an example of a geometric series. We discuss geometric series in more detail later in this section. First, we summarize what it means for an infinite series to converge.

Definition

An **infinite series** is an expression of the form

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots.$$

For each positive integer k, the sum

$$S_k = \sum_{n=1}^{k} a_n = a_1 + a_2 + a_3 + \cdots + a_k$$

is called the kth **partial sum** of the infinite series. The partial sums form a sequence S_k. If the sequence of partial sums converges to a real number S, the infinite series converges. If we can describe the **convergence of a series** to S, we call S the sum of the series, and we write

$$\sum_{n=1}^{\infty} a_n = S.$$

If the sequence of partial sums diverges, we have the **divergence of a series.**

Note that the index for a series need not begin with $n = 1$ but can begin with any value. For example, the series

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1}$$

can also be written as

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n} \text{ or } \sum_{n=5}^{\infty} \left(\frac{1}{2}\right)^{n-5}.$$

Often it is convenient for the index to begin at 1, so if for some reason it begins at a different value, we can re-index by making a change of variables. For example, consider the series

$$\sum_{n=2}^{\infty} \frac{1}{n^2}.$$

By introducing the variable $m = n - 1$, so that $n = m + 1$, we can rewrite the series as

$$\sum_{m=1}^{\infty} \frac{1}{(m+1)^2}.$$

Example 5.2.1: Evaluating Limits of Sequences of Partial Sums

For each of the following series, use the sequence of partial sums to determine whether the series converges or diverges.

a. $\displaystyle\sum_{n=1}^{\infty} \frac{n}{n+1}$

b. $\displaystyle\sum_{n=1}^{\infty} (-1)^n$

c. $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$

Solution

a. The sequence of partial sums S_k satisfies

$$S_1 = \frac{1}{2}$$
$$S_2 = \frac{1}{2} + \frac{2}{3}$$
$$S_3 = \frac{1}{2} + \frac{2}{3} + \frac{3}{4}$$
$$S_4 = \frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \frac{4}{5} \quad .$$

Notice that each term added is greater than $1/2$. As a result, we see that

$$S_1 = \frac{1}{2}$$

$$S_2 = \frac{1}{2} + \frac{2}{3} > \frac{1}{2} + \frac{1}{2} = 2\left(\frac{1}{2}\right)$$
$$S_3 = \frac{1}{2} + \frac{2}{3} + \frac{3}{4} > \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 3\left(\frac{1}{2}\right)$$
$$S_4 = \frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \frac{4}{5} > \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 4\left(\frac{1}{2}\right).$$

From this pattern we can see that $S_k > k\left(\frac{1}{2}\right)$ for every integer k. Therefore, S_k is unbounded and consequently, diverges. Therefore, the infinite series $\sum_{n=1}^{\infty} \frac{n}{n+1}$ diverges.

b. The sequence of partial sums S_k satisfies

$$S_1 = -1$$
$$S_2 = -1 + 1 = 0$$
$$S_3 = -1 + 1 - 1 = -1$$
$$S_4 = -1 + 1 - 1 + 1 = 0.$$

From this pattern we can see the sequence of partial sums is

$$S_k = -1, 0, -1, 0, \ldots.$$

Since this sequence diverges, the infinite series $\sum_{n=1}^{\infty} (-1)^n$ diverges.

c. The sequence of partial sums S_k satisfies

$$S_1 = \frac{1}{1\cdot 2} = \frac{1}{2}$$
$$S_2 = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$
$$S_3 = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} = \frac{3}{4}$$
$$S_4 = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} = \frac{4}{5}$$
$$S_5 = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} + \frac{1}{5\cdot 6} = \frac{5}{6}.$$

From this pattern, we can see that the kth partial sum is given by the explicit formula

$$S_k = \frac{k}{k+1}$$

.

Since $k/(k+1) \to 1,$ we conclude that the sequence of partial sums converges, and therefore the infinite series converges to 1. We have

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

Exercise 5.2.1

Determine whether the series $\sum_{n=1}^{\infty} \frac{n+1}{n}$ converges or diverges.

Hint

Look at the sequence of partial sums.

Answer

The series diverges because the kth partial sum $S_k > k$.

The Harmonic Series

A useful series to know about is the **harmonic series.** The harmonic series is defined as

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

This series is interesting because it diverges, but it diverges very slowly. By this we mean that the terms in the sequence of partial sums S_k approach infinity, but do so very slowly. We will show that the series diverges, but first we illustrate the slow growth of the terms in the sequence S_k in the following table.

k	10	100	1000	10,00	100,000	1,000,000
S_k	2.92897	5.18738	7.48547	9.78761	12.09015	14.39273

Even after $1,000,000$terms, the partial sum is still relatively small. From this table, it is not clear that this series actually diverges. However, we can show analytically that the sequence of partial sums diverges, and therefore the series diverges.

To show that the sequence of partial sums diverges, we show that the sequence of partial sums is unbounded. We begin by writing the first several partial sums:

$$\begin{aligned} S_1 &= 1 \\ S_2 &= 1 + \frac{1}{2} \\ S_3 &= 1 + \frac{1}{2} + \frac{1}{3} \\ S_4 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \quad . \end{aligned}$$

Notice that for the last two terms in S_4,

$$\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4}$$

Therefore, we conclude that

$$S_4 > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 1 + \frac{1}{2} + \frac{1}{2} = 1 + 2\left(\frac{1}{2}\right).$$

Using the same idea for S_8, we see that

$$S_8 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1 + 3\left(\frac{1}{2}\right).$$

From this pattern, we see that $S_1 = 1, S_2 = 1 + 1/2, S_4 > 1 + 2(1/2),$ and $S_8 > 1 + 3(1/2)$. More generally, it can be shown that $S_{2^j} > 1 + j(1/2)$ for all $j > 1$. Since $1 + j(1/2) \to \infty,$ we conclude that the sequence S_k is unbounded and therefore diverges. In the previous section, we stated that convergent sequences are bounded. Consequently, since S_k is unbounded, it diverges. Thus, the harmonic series diverges.

Algebraic Properties of Convergent Series

Since the sum of a convergent infinite series is defined as a limit of a sequence, the algebraic properties for series listed below follow directly from the algebraic properties for sequences.

Note 5.2.1: Algebraic Properties of Convergent Series

Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be convergent series. Then the following algebraic properties hold.

i. The series $\sum_{n=1}^{\infty}(a_n+b_n)$ converges, and $\sum_{n=1}^{\infty}(a_n+b_n)=\sum_{n=1}^{\infty}a_n+\sum_{n=1}^{\infty}b_n$. (Sum Rule)

ii. The series $\sum_{n=1}^{\infty}(a_n-b_n)$ converges, and $\sum_{n=1}^{\infty}(a_n-b_n)=\sum_{n=1}^{\infty}a_n-\sum_{n=1}^{\infty}b_n$. (Difference Rule)

iii. For any real number c, the series $\sum_{n=1}^{\infty} ca_n$ converges, and $\sum_{n=1}^{\infty} ca_n = c\sum_{n=1}^{\infty} a_n$. (Constant Multiple Rule)

Example 5.2.2: Using Algebraic Properties of Convergent Series

Evaluate $\sum_{n=1}^{\infty}\left[\frac{3}{n(n+1)}+\left(\frac{1}{2}\right)^{n-2}\right]$.

Solution

We showed earlier that

$$\sum_{n=1}^{\infty}\frac{1}{n(n+1)}=1$$

and

$$\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^{n-1}=2.$$

Since both of those series converge, we can apply the properties of Note 5.2.1 to evaluate

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

Using the sum rule, write

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

Then, using the constant multiple rule and the sums above, we can conclude that

$$\sum_{n=1}^{\infty}\frac{3}{n(n+1)}+\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^{n-2}=3\sum_{n=1}^{\infty}\frac{1}{n(n+1)}+\left(\frac{1}{2}\right)^{-1}\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^{n-1}=3(1)+\left(\frac{1}{2}\right)^{-1}(2)=3+2(2)=7.$$

Exercise 5.2.2

Evaluate $\sum_{n=1}^{\infty}\frac{5}{2^{n-1}}$.

Hint

Rewrite as $\sum_{n=1}^{\infty}5\left(\frac{1}{2}\right)^{n-1}$.

Answer

10

Geometric Series

A **geometric series** is any series that we can write in the form

$$a + ar + ar^2 + ar^3 + \cdots = \sum_{n=1}^{\infty} ar^{n-1}.$$

Because the ratio of each term in this series to the previous term is r, the number r is called the ratio. We refer to a as the initial term because it is the first term in the series. For example, the series

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

is a geometric series with initial term $a = 1$ and ratio $r = 1/2$.

In general, when does a geometric series converge? Consider the geometric series

$$\sum_{n=1}^{\infty} ar^{n-1}$$

when $a > 0$. Its sequence of partial sums S_k is given by

$$S_k = \sum_{n=1}^{k} ar^{n-1} = a + ar + ar^2 + \cdots + ar^{k-1}.$$

Consider the case when $r = 1$. In that case,

$$S_k = a + a(1) + a(1)^2 + \cdots + a(1)^{k-1} = ak.$$

Since $a > 0$, we know $ak \to \infty$ as $k \to \infty$. Therefore, the sequence of partial sums is unbounded and thus diverges. Consequently, the infinite series diverges for $r = 1$. For $r \neq 1$, to find the limit of S_k, multiply Equation by $1 - r$. Doing so, we see that

$$\begin{aligned}(1-r)S_k = a(1-r)(1 + r + r^2 + r^3 + \cdots + r^{k-1}) &= a[(1 + r + r^2 + r^3 + \cdots + r^{k-1}) - (r + r^2 + r^3 + \cdots + r^k)] \\ &= a(1 - r^k).\end{aligned}$$

All the other terms cancel out.

Therefore,

$$S_k = \frac{a(1-r^k)}{1-r} \text{ for } r \neq 1.$$

From our discussion in the previous section, we know that the geometric sequence $r^k \to 0$ if $|r| < 1$ and that r^k diverges if $|r| > 1$ or $r = \pm 1$. Therefore, for $|r| < 1$, $S_k \to \dfrac{a}{1-r}$ and we have

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r} \text{ if } |r| < 1.$$

If $|r| \geq 1$, S_k diverges, and therefore

$$\sum_{n=1}^{\infty} ar^{n-1} \text{ diverges if } |r| \geq 1.$$

Definitions: Diverging and Converging Series

A geometric series is a series of the form

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + ar^3 + \cdots. \tag{5.2.1}$$

If $|r| < 1$, the series **converges**, and

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r} \text{ for } |r| < 1.$$

If $|r| \geq 1$, the series **diverges**.

Geometric series sometimes appear in slightly different forms. For example, sometimes the index begins at a value other than $n = 1$ or the exponent involves a linear expression for n other than $n - 1$. As long as we can rewrite the series in the form given by Equation 5.2.1, it is a geometric series. For example, consider the series

$$\sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^{n+2}.$$

To see that this is a geometric series, we write out the first several terms:

$$\sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^{n+2} = \left(\frac{2}{3}\right)^{2} + \left(\frac{2}{3}\right)^{3} + \left(\frac{2}{3}\right)^{4} + \cdots = \frac{4}{9} + \frac{4}{9} \cdot \left(\frac{2}{3}\right) + \frac{4}{9} \cdot \left(\frac{2}{3}\right)^{2} + \cdots.$$

We see that the initial term is $a = 4/9$ and the ratio is $r = 2/3$. Therefore, the series can be written as

$$\sum_{n=1}^{\infty} \frac{4}{9} \cdot \left(\frac{2}{3}\right)^{n-1}.$$

Since $r = 2/3 < 1$, this series converges, and its sum is given by

$$\sum_{n=1}^{\infty} \frac{4}{9} \cdot \left(\frac{2}{3}\right)^{n-1} = \frac{4/9}{1 - 2/3} = \frac{4}{3}.$$

Example 5.2.3: Determining Convergence or Divergence of a Geometric Series

Determine whether each of the following geometric series converges or diverges, and if it converges, find its sum.

a. $\displaystyle\sum_{n=1}^{\infty} \frac{(-3)^{n+1}}{4^{n-1}}$

b. $\displaystyle\sum_{n=1}^{\infty} e^{2n}$

Solution

a. Writing out the first several terms in the series, we have

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{(-3)^{n+1}}{4^{n-1}} &= \frac{(-3)^2}{4^0} + \frac{(-3)^3}{4} + \frac{(-3)^4}{4^2} + \cdots \\ &= (-3)^2 + (-3)^2 \cdot \left(\frac{-3}{4}\right) + (-3)^2 \cdot \left(\frac{-3}{4}\right)^2 + \cdots \\ &= 9 + 9 \cdot \left(\frac{-3}{4}\right) + 9 \cdot \left(\frac{-3}{4}\right)^2 + \cdots.\end{aligned}$$

The initial term $a = -3$ and the ratio $r = -3/4$. Since $|r| = 3/4 < 1$, the series converges to

$$\frac{9}{1 - (-3/4)} = \frac{9}{7/4} = \frac{36}{7}.$$

b. Writing this series as

$$e^2 \sum_{n=1}^{\infty} (e^2)^{n-1}$$

we can see that this is a geometric series where $r = e^2 > 1$. Therefore, the series diverges.

Exercise 5.2.3

Determine whether the series $\sum_{n=1}^{\infty} \left(\frac{-2}{5}\right)^{n-1}$ converges or diverges. If it converges, find its sum.

Hint

$r = -2/5$

Answer

$5/7$

We now turn our attention to a nice application of geometric series. We show how they can be used to write repeating decimals as fractions of integers.

Example 5.2.4: Writing Repeating Decimals as Fractions of Integers

Use a geometric series to write $3.2\bar{6}$ as a fraction of integers.

Solution

Since $3.2\bar{6} = 3.262626\ldots$, first we write

$$\begin{aligned} 3.262626\ldots &= 3 + \frac{26}{100} + \frac{26}{10,000} + \frac{26}{100,000} + \cdots \\ &= 3 + \frac{26}{10^2} + \frac{26}{10^4} + \frac{26}{10^6} + \cdots. \end{aligned}$$

Ignoring the term 3, the rest of this expression is a geometric series with initial term $a = 26/10^2$ and ratio $r = 1/10^2$. Therefore, the sum of this series is

$$\frac{26/10^2}{1-(1/10^2)} = \frac{26/10^2}{99/10^2} = \frac{26}{99}.$$

Thus,

$$3.262626\ldots = 3 + \tfrac{26}{99} = \tfrac{323}{99}.$$

Exercise 5.2.4

Write $5.2\bar{7}$ as a fraction of integers.

Hint

By expressing this number as a series, find a geometric series with initial term $a = 7/100$ and ratio $r = 1/10$.

Answer

$475/90$

Example 5.2.5: Finding the Area of the Koch Snowflake

Define a sequence of figures $\{F_n\}$ recursively as follows (Figure 5.2.2). Let F_0 be an equilateral triangle with sides of length 1. For $n \geq 1$, let F_n be the curve created by removing the middle third of each side of F_{n-1} and replacing it with an equilateral triangle pointing outward. The limiting figure as $n \to \infty$ is known as **Koch's snowflake.**

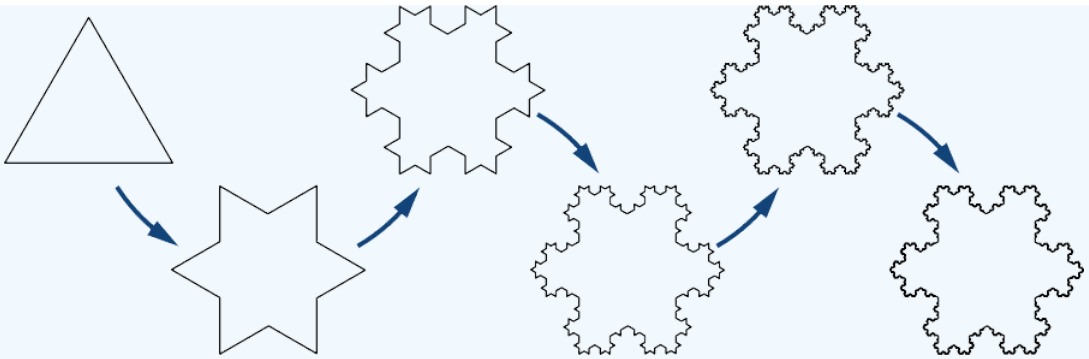


Figure 5.2.2: The first four figures, F_0, F_1, F_2, and F_3, in the construction of the Koch snowflake.

a. Find the length L_n of the perimeter of F_n. Evaluate $\lim_{n\to\infty} L_n$ to find the length of the perimeter of Koch's snowflake.

b. Find the area A_n of figure F_n. Evaluate $\lim_{n\to\infty} A_n$ to find the area of Koch's snowflake.

Solution

a. Let N_n denote the number of sides of figure F_n. Since F_0 is a triangle, $N_0 = 3$. Let l_n denote the length of each side of F_n. Since F_0 is an equilateral triangle with sides of length $l_0 = 1$, we now need to determine N_1 and l_1. Since F_1 is created by removing the middle third of each side and replacing that line segment with two line segments, for each side of F_0, we get four sides in F_1. Therefore, the number of sides for F_1 is

$$N_1 = 4 \cdot 3\,.$$

Since the length of each of these new line segments is $1/3$ the length of the line segments in F_0, the length of the line segments for F_1 is given by

$$l_1 = \tfrac{1}{3} \cdot 1 = \tfrac{1}{3}\ .$$

Similarly, for F_2, since the middle third of each side of F_1 is removed and replaced with two line segments, the number of sides in F_2 is given by

$$N_2 = 4N_1 = 4(4 \cdot 3) = 4^2 \cdot 3.$$

Since the length of each of these sides is $1/3$ the length of the sides of F_1, the length of each side of figure F_2 is given by

$$l_2 = \tfrac{1}{3} \cdot l_1 = \tfrac{1}{3} \cdot \tfrac{1}{3} = \left(\tfrac{1}{3}\right)^2\ .$$

More generally, since F_n is created by removing the middle third of each side of F_{n-1} and replacing that line segment with two line segments of length $\frac{1}{3}l_{n-1}$ in the shape of an equilateral triangle, we know that $N_n = 4N_{n-1}$ and $l_n = \dfrac{l_{n-1}}{3}$. Therefore, the number of sides of figure F_n is

$$N_n = 4^n \cdot 3$$

and the length of each side is

$$l_n = \left(\frac{1}{3}\right)^n.$$

Therefore, to calculate the perimeter of F_n, we multiply the number of sides N_n and the length of each side l_n. We conclude that the perimeter of F_n is given by

$$L_n = N_n \cdot l_n = 3 \cdot \left(\frac{4}{3}\right)^n$$

Therefore, the length of the perimeter of Koch's snowflake is

$$L = \lim_{n\to\infty} L_n = \infty.$$

b. Let T_n denote the area of each new triangle created when forming F_n. For $n = 0, T_0$ is the area of the original equilateral triangle. Therefore, $T_0 = A_0 = \sqrt{3}/4$. For $n \geq 1$, since the lengths of the sides of the new triangle are $1/3$ the length of the

sides of F_{n-1}, we have

$$T_n = \left(\frac{1}{3}\right)^2 \cdot T_{n-1} = \frac{1}{9} \cdot T_{n-1}.$$

Therefore, $T_n = \left(\frac{1}{9}\right)^n \cdot \frac{\sqrt{3}}{4}$. Since a new triangle is formed on each side of F_{n-1},

$$A_n = A_{n-1} + N_{n-1} \cdot T_n = A_{n-1} + (3 \cdot 4_{n-1}) \cdot \left(\frac{1}{9}\right)^n \cdot \frac{\sqrt{3}}{4} = A_{n-1} + \frac{3}{4} \cdot \left(\frac{4}{9}\right)^n \cdot \frac{\sqrt{3}}{4}.$$

Writing out the first few terms A_0, A_1, A_2, we see that

$$A_0 = \tfrac{\sqrt{3}}{4}$$

$$A_1 = A_0 + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right) \cdot \tfrac{\sqrt{3}}{4} = \tfrac{\sqrt{3}}{4} + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right) \cdot \tfrac{\sqrt{3}}{4} = \tfrac{\sqrt{3}}{4}\left[1 + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right)\right]$$

$$A_2 = A_1 + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right)^2 \cdot \tfrac{\sqrt{3}}{4} = \tfrac{\sqrt{3}}{4}\left[1 + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right)\right] + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right)^2 \cdot \tfrac{\sqrt{3}}{4} = \tfrac{\sqrt{3}}{4}\left[1 + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right) + \tfrac{3}{4} \cdot \left(\tfrac{4}{9}\right)^2\right] \quad .$$

More generally,

$$A_n = \tfrac{\sqrt{3}}{4}\left[1 + \tfrac{3}{4}\left(\tfrac{4}{9} + \left(\tfrac{4}{9}\right)^2 + \cdots + \left(\tfrac{4}{9}\right)^n\right)\right] \quad .$$

Factoring $4/9$ out of each term inside the inner parentheses, we rewrite our expression as

$$A_n = \tfrac{\sqrt{3}}{4}\left[1 + \tfrac{1}{3}\left(1 + \tfrac{4}{9} + \left(\tfrac{4}{9}\right)^2 + \cdots + \left(\tfrac{4}{9}\right)^{n-1}\right)\right] \quad .$$

The expression $1 + \left(\frac{4}{9}\right) + \left(\frac{4}{9}\right)^2 + \cdots + \left(\frac{4}{9}\right)^{n-1}$ is a geometric sum. As shown earlier, this sum satisfies

$$1 + \tfrac{4}{9} + \left(\tfrac{4}{9}\right)^2 + \cdots + \left(\tfrac{4}{9}\right)^{n-1} = \frac{1 - (4/9)^n}{1 - (4/9)}.$$

Substituting this expression into the expression above and simplifying, we conclude that

$$A_n = \frac{\sqrt{3}}{4}\left[1 + \frac{1}{3}\left(\frac{1 - (4/9)^n}{1 - (4/9)}\right)\right] = \frac{\sqrt{3}}{4}\left[\frac{8}{5} - \frac{3}{5}\left(\frac{4}{9}\right)^n\right].$$

Therefore, the area of Koch's snowflake is

$$A = \lim_{n \to \infty} A_n = \frac{2\sqrt{3}}{5}.$$

Analysis

The Koch snowflake is interesting because it has finite area, yet infinite perimeter. Although at first this may seem impossible, recall that you have seen similar examples earlier in the text. For example, consider the region bounded by the curve $y = 1/x^2$ and the x-axis on the interval $[1, \infty)$. Since the improper integral

$$\int_1^\infty \frac{1}{x^2}\, dx$$

converges, the area of this region is finite, even though the perimeter is infinite.

Telescoping Series

Consider the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$. We discussed this series in Example 5.2.1c, showing that the series converges by writing out the first several partial sums $S_1, S_2, \ldots, S_6$ and noticing that they are all of the form $S_k = \dfrac{k}{k+1}$. Here we use a different technique to show that this series converges. By using partial fractions, we can write

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

Therefore, the series can be written as

$$\sum_{n=1}^{\infty}\left[\frac{1}{n}-\frac{1}{n+1}\right]=\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\cdots.$$

Writing out the first several terms in the sequence of partial sums S_k, we see that

$$\begin{aligned}S_1&=1-\tfrac{1}{2}\\S_2&=\left(1-\tfrac{1}{2}\right)+\left(\tfrac{1}{2}-\tfrac{1}{3}\right)=1-\tfrac{1}{3}\\S_3&=\left(1-\tfrac{1}{2}\right)+\left(\tfrac{1}{2}-\tfrac{1}{3}\right)+\left(\tfrac{1}{3}-\tfrac{1}{4}\right)=1-\tfrac{1}{4}\quad.\end{aligned}$$

In general,

$$S_k=\left(1-\tfrac{1}{2}\right)+\left(\tfrac{1}{2}-\tfrac{1}{3}\right)+\left(\tfrac{1}{3}-\tfrac{1}{4}\right)+\cdots+\left(\tfrac{1}{k}-\tfrac{1}{k+1}\right)=1-\frac{1}{k+1}\quad.$$

We notice that the middle terms cancel each other out, leaving only the first and last terms. In a sense, the series collapses like a spyglass with tubes that disappear into each other to shorten the telescope. For this reason, we call a series that has this property a telescoping series. For this series, since $S_k=1-1/(k+1)$ and $1/(k+1)\to 0$ as $k\to\infty$, the sequence of partial sums converges to 1, and therefore the series converges to 1.

Definition

A **telescoping series** is a series in which most of the terms cancel in each of the partial sums, leaving only some of the first terms and some of the last terms.

For example, any series of the form

$$\sum_{n=1}^{\infty}[b_n-b_{n+1}]=(b_1-b_2)+(b_2-b_3)+(b_3-b_4)+\cdots$$

is a telescoping series. We can see this by writing out some of the partial sums. In particular, we see that

$$\begin{aligned}S_1&=b_1-b_2\\S_2&=(b_1-b_2)+(b_2-b_3)=b_1-b_3\\S_3&=(b_1-b_2)+(b_2-b_3)+(b_3-b_4)=b_1-b_4.\end{aligned}$$

In general, the kth partial sum of this series is

$$S_k=b_1-b_{k+1}\ .$$

Since the kth partial sum can be simplified to the difference of these two terms, the sequence of partial sums S_k will converge if and only if the sequence b_{k+1} converges. Moreover, if the sequence b_{k+1} converges to some finite number B, then the sequence of partial sums converges to b_1-B, and therefore

$$\sum_{n=1}^{\infty}[b_n-b_{n+1}]=b_1-B.$$

In the next example, we show how to use these ideas to analyze a telescoping series of this form.

Example 5.2.6: Evaluating a Telescoping Series

Determine whether the telescoping series

$$\sum_{n=1}^{\infty}\left[\cos\left(\frac{1}{n}\right)-\cos\left(\frac{1}{n+1}\right)\right]$$

converges or diverges. If it converges, find its sum.

Solution

By writing out terms in the sequence of partial sums, we can see that

$$S_1=\cos(1)-\cos(\tfrac{1}{2})$$

$S_2 = (\cos(1) - \cos(\frac{1}{2})) + (\cos(\frac{1}{2}) - \cos(\frac{1}{3})) = \cos(1) - \cos(\frac{1}{3})$

$S_3 = (\cos(1) - \cos(\frac{1}{2})) + (\cos(\frac{1}{2}) - \cos(\frac{1}{3})) + (\cos(\frac{1}{3}) - \cos(\frac{1}{4}))$

$= \cos(1) - \cos(\frac{1}{4})$.

In general,

$$S_k = \cos(1) - \cos\left(\tfrac{1}{k+1}\right).$$

Since $1/(k+1) \to 0$ as $k \to \infty$ and $\cos x$ is a continuous function, $\cos(1/(k+1)) \to \cos(0) = 1$. Therefore, we conclude that $S_k \to \cos(1) - 1$. The telescoping series converges and the sum is given by

$$\sum_{n=1}^{\infty} \left[\cos\left(\frac{1}{n}\right) - \cos\left(\frac{1}{n+1}\right)\right] = \cos(1) - 1.$$

Exercise 5.2.5

Determine whether $\sum_{n=1}^{\infty} [e^{1/n} - e^{1/(n+1)}]$ converges or diverges. If it converges, find its sum.

Hint

Write out the sequence of partial sums to see which terms cancel.

Answer

$e - 1$

Euler's Constant

We have shown that the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. Here we investigate the behavior of the partial sums S_k as $k \to \infty$. In particular, we show that they behave like the natural logarithm function by showing that there exists a constant γ such that

$$\sum_{n=1}^{k} \left(\frac{1}{n} - \ln k\right) \to \gamma \text{ as } k \to \infty.$$

This constant γ is known as **Euler's constant.**

1. Let $T_k = \sum_{n=1}^{k} \left(\frac{1}{n} - \ln k\right)$. Evaluate T_k for various values of k.

2. For T_k as defined in part 1. show that the sequence T_k converges by using the following steps.

   a. Show that the sequence T_k is monotone decreasing. (Hint: Show that $\ln(1 + 1/k > 1/(k+1))$

   b. Show that the sequence T_k is bounded below by zero. (Hint: Express $\ln k$ as a definite integral.)

   c. Use the Monotone Convergence Theorem to conclude that the sequence T_k converges. The limit γ is Euler's constant.

3. Now estimate how far T_k is from γ for a given integer k. Prove that for $k \geq 1, 0 < T_k - \gamma \leq 1/k$ by using the following steps.

   a. Show that $\ln(k+1) - \ln k < 1/k$.

   b. Use the result from part a. to show that for any integer k,

   $$T_k - T_{k+1} < \frac{1}{k} - \frac{1}{k+1}.$$

   c. For any integers k and j such that $j > k$, express $T_k - T_j$ as a telescoping sum by writing

$$T_k - T_j = (T_k - T_{k+1}) + (T_{k+1} - T_{k+2}) + (T_{k+2} - T_{k+3}) + \cdots + (T_{j-1} - T_j).$$

Use the result from part b. combined with this telescoping sum to conclude that

$$T_k - T_j < \frac{1}{k} - \frac{1}{j}.$$

a. Apply the limit to both sides of the inequality in part c. to conclude that

$$T_k - \gamma \leq \frac{1}{k}.$$

e. Estimate γ to an accuracy of within **0.001**.

Key Concepts

- Given the infinite series

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots$$

and the corresponding sequence of partial sums S_k where

$$S_k = \sum_{n=1}^{k} a_n = a_1 + a_2 + a_3 + \cdots + a_k \quad ,$$

the series converges if and only if the sequence S_k converges.

- The geometric series $\sum_{n=1}^{\infty} ar^{n-1}$ converges if $|r| < 1$ and diverges if $|r| \geq 1$. For $|r| < 1$,

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r} \, .$$

- The harmonic series

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

diverges.

- A series of the form $\sum_{n=1}^{\infty} [b_n - b_{n+1}] = [b_1 - b_2] + [b_2 - b_3] + [b_3 - b_4] + \cdots + [b_n - b_{n+1}] + \cdots$ is a telescoping series. The kth partial sum of this series is given by $S_k = b_1 - b_{k+1}$. The series will converge if and only if $\lim_{k \to \infty} b_{k+1}$ exists. In that case,

$$\sum_{n=1}^{\infty} [b_n - b_{n+1}] = b_1 - \lim_{k \to \infty} (b_{k+1}) \, .$$

Key Equations

- **Harmonic series**

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

- **Sum of a geometric series**

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r} \quad \text{for } |r| < 1$$

Glossary

convergence of a series

a series converges if the sequence of partial sums for that series converges

divergence of a series

a series diverges if the sequence of partial sums for that series diverges

geometric series

a geometric series is a series that can be written in the form

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + ar^3 + \cdots$$

harmonic series

the harmonic series takes the form

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

infinite series

an infinite series is an expression of the form

$$a_1 + a_2 + a_3 + \cdots = \sum_{n=1}^{\infty} a_n$$

partial sum

the kth partial sum of the infinite series $\sum_{n=1}^{\infty} a_n$ is the finite sum

$$S_k = \sum_{n=1}^{k} a_n = a_1 + a_2 + a_3 + \cdots + a_k$$

telescoping series

a telescoping series is one in which most of the terms cancel in each of the partial sums

---

This page titled 5.2: Infinite Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.2: Infinite Series** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

5.3: The Divergence and Integral Tests

Learning Objectives

- Use the divergence test to determine whether a series converges or diverges.
- Use the integral test to determine the convergence of a series.
- Estimate the value of a series by finding bounds on its remainder term.

In the previous section, we determined the convergence or divergence of several series by explicitly calculating the limit of the sequence of partial sums S_k. In practice, explicitly calculating this limit can be difficult or impossible. Luckily, several tests exist that allow us to determine convergence or divergence for many types of series. In this section, we discuss two of these tests: the divergence test and the integral test. We will examine several other tests in the rest of this chapter and then summarize how and when to use them.

Divergence Test

For a series $\sum_{n=1}^{\infty} a_n$ to converge, the nth term a_n must satisfy $a_n \to 0$ as $n \to \infty$. Therefore, from the algebraic limit properties of sequences,

$$\begin{aligned} \lim_{k\to\infty} a_k &= \lim_{k\to\infty} (S_k - S_{k-1}) \\ &= \lim_{k\to\infty} S_k - \lim_{k\to\infty} S_{k-1} \\ &= S - S = 0. \end{aligned}$$

Therefore, if $\sum_{n=1}^{\infty} a_n$ converges, the nth term $a_n \to 0$ as $n \to \infty$. An important consequence of this fact is the following statement:

$$\text{If } a_n \nrightarrow 0 \text{ as } n \to \infty, \sum_{n=1}^{\infty} a_n \text{ diverges.}$$

This test is known as the divergence test because it provides a way of proving that a series diverges.

Definition: The Divergence Test

If $\lim_{n\to\infty} a_n = c \neq 0$ or $\lim_{n\to\infty} a_n$ does not exist, then the series $\sum_{n=1}^{\infty} a_n$ diverges.

It is important to note that the inverse statement of this theorem is not true. That is, if $\lim_{n\to\infty} a_n = 0$, we cannot make any conclusion about the convergence of $\sum_{n=1}^{\infty} a_n$.

For example, $\lim_{n\to 0} \frac{1}{n} = 0$, but the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges. In this section and the remaining sections of this chapter, we show many more examples of such series. Consequently, although we can use the divergence test to show that a series diverges, we cannot use it to prove that a series converges. Specifically, if $a_n \to 0$, the divergence test is inconclusive.

Example 5.3.1: Using the divergence test

For each of the following series, apply the divergence test. If the divergence test proves that the series diverges, state so. Otherwise, indicate that the divergence test is inconclusive.

a. $\sum_{n=1}^{\infty} \frac{n}{3n-1}$

b. $\sum_{n=1}^{\infty} \frac{1}{n^3}$

c. $\sum_{n=1}^{\infty} e^{1/n^2}$

Solution

a. Since $\lim_{n\to\infty} \frac{n}{3n-1} = \frac{1}{3} \neq 0$, by the divergence test, we can conclude that $\sum_{n=1}^{\infty} \frac{n}{3n-1}$ diverges.

b. Since $\lim_{n\to\infty} \frac{1}{n^3} = 0$, the divergence test is inconclusive.

c. Since $\lim_{n\to\infty} e^{1/n^2} = 1 \neq 0$, by the divergence test, the series $\sum_{n=1}^{\infty} e^{1/n^2}$ diverges.

? Exercise 5.3.1

What does the divergence test tell us about the series $\sum_{n=1}^{\infty} \cos(1/n^2)$?

Hint

Look at $\lim_{n\to\infty} \cos(1/n^2)$.

Answer

The series diverges.

Integral Test

In the previous section, we proved that the harmonic series diverges by looking at the sequence of partial sums S_k and showing that $S_{2^k} > 1 + k/2$ for all positive integers k. In this section we use a different technique to prove the divergence of the harmonic series. This technique is important because it is used to prove the divergence or convergence of many other series. This test, called the **integral test,** compares an infinite sum to an improper integral. It is important to note that this test can only be applied when we are considering a series whose terms are all positive.

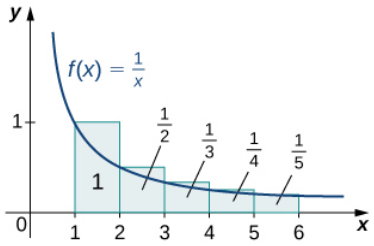


Figure 5.3.1: *The sum of the areas of the rectangles is greater than the area between the curve $f(x) = 1/x$ and the x-axis for $x \geq 1$. Since the area bounded by the curve is infinite (as calculated by an improper integral), the sum of the areas of the rectangles is also infinite.*

To illustrate how the integral test works, use the harmonic series as an example. In Figure 5.3.1, we depict the harmonic series by sketching a sequence of rectangles with areas $1, 1/2, 1/3, 1/4, \ldots$ along with the function $f(x) = 1/x$. From the graph, we see that

$$\sum_{n=1}^{k} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k} > \int_1^{k+1} \frac{1}{x}\, dx.$$

Therefore, for each k, the kth partial sum S_k satisfies

$$S_k = \sum_{n=1}^{k} \frac{1}{n} > \int_1^{k+1} \frac{1}{x}\,dx = \ln x\Big|_1^{k+1}$$
$$= \ln(k+1) - \ln(1)$$
$$= \ln(k+1).$$

Since $\lim_{k\to\infty} \ln(k+1) = \infty$, we see that the sequence of partial sums S_k is unbounded. Therefore, S_k diverges, and, consequently, the series $\sum_{n=1}^{\infty} \frac{1}{n}$ also diverges.

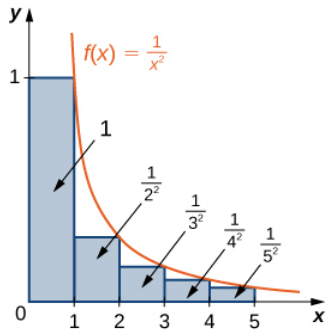


Figure 5.3.2: The sum of the areas of the rectangles is less than the sum of the area of the first rectangle and the area between the curve $f(x) = 1/x^2$ and the x-axis for $x \geq 1$. Since the area bounded by the curve is finite, the sum of the areas of the rectangles is also finite.

Now consider the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$. We show how an integral can be used to prove that this series converges. In Figure 5.3.2, we sketch a sequence of rectangles with areas $1, 1/2^2, 1/3^2, \ldots$ along with the function $f(x) = \frac{1}{x^2}$. From the graph we see that

$$\sum_{n=1}^{k} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{k^2} < 1 + \int_1^k \frac{1}{x^2}\,dx.$$

Therefore, for each k, the kth partial sum S_k satisfies

$$S_k = \sum_{n=1}^{k} \frac{1}{n^2} < 1 + \int_1^k \frac{1}{x^2}\,dx = 1 - \frac{1}{x}\Big|_1^k$$
$$= 1 - \frac{1}{k} + 1$$
$$= 2 - \frac{1}{k} < 2.$$

We conclude that the sequence of partial sums S_k is bounded. We also see that S_k is an increasing sequence:

$$S_k = S_{k-1} + \frac{1}{k^2}$$

for $k \geq 2$.

Since S_k is increasing and bounded, by the **Monotone Convergence Theorem**, it converges. Therefore, the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

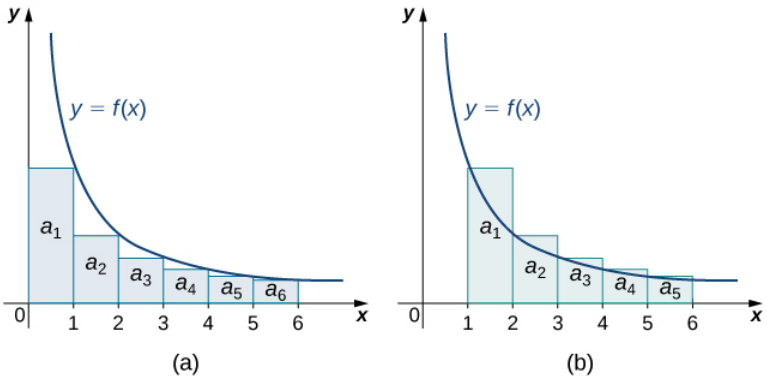


Figure 5.3.3: (a) If we can inscribe rectangles inside a region bounded by a curve $y = f(x)$ and the x-axis, and the area bounded by those curves for $x \geq 1$ is finite, then the sum of the areas of the rectangles is also finite. (b) If a set of rectangles circumscribes the region bounded by $y = f(x)$ and the x axis for $x \geq 1$ and the region has infinite area, then the sum of the areas of the rectangles is also infinite.

We can extend this idea to prove convergence or divergence for many different series. Suppose $\sum_{n=1}^{\infty} a_n$ is a series with positive terms a_n such that there exists a continuous, positive, decreasing function f where $f(n) = a_n$ for all positive integers. Then, as in Figure 5.3.3*a*, for any integer k, the kth partial sum S_k satisfies

$$S_k = a_1 + a_2 + a_3 + \cdots + a_k < a_1 + \int_1^k f(x)\,dx < 1 + \int_1^{\infty} f(x)\,dx.$$

Therefore, if $\int_1^{\infty} f(x)\,dx$ converges, then the sequence of partial sums S_k is bounded. Since S_k is an increasing sequence, if it is also a bounded sequence, then by the Monotone Convergence Theorem, it converges. We conclude that if $\int_1^{\infty} f(x)\,dx$ converges, then the series $\sum_{n=1}^{\infty} a_n$ also converges. On the other hand, from Figure 5.3.3*b*, for any integer k, the kth partial sum S_k satisfies

$$S_k = a_1 + a_2 + a_3 + \cdots + a_k > \int_1^{k+1} f(x)\,dx.$$

If

$$\lim_{k\to\infty} \int_1^{k+1} f(x)\,dx = \infty,$$

then S_k is an unbounded sequence and therefore diverges. As a result, the series $\sum_{n=1}^{\infty} a_n$ also diverges. Since f is a positive function, if $\int_1^{\infty} f(x)\,dx$ diverges, then

$$\lim_{k\to\infty} \int_1^{k+1} f(x)\,dx = \infty.$$

We conclude that if $\int_1^{\infty} f(x)\,dx$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

Definition: The Integral Test

Suppose $\sum_{n=1}^{\infty} a_n$ is a series with positive terms a_n. Suppose there exists a function f and a positive integer N such that the following three conditions are satisfied:

i. f is continuous,
ii. f is decreasing, and

iii. $f(n) = a_n$ for all integers $n \geq N$.

Then

$$\sum_{n=1}^{\infty} a_n$$

and

$$\int_N^{\infty} f(x)\,dx$$

both converge or both diverge (Figure 5.3.3).

Although convergence of $\int_N^{\infty} f(x)\,dx$ implies convergence of the related series $\sum_{n=1}^{\infty} a_n$, it does not imply that the value of the integral and the series are the same. They may be different, and often are. For example,

$$\sum_{n=1}^{\infty} \left(\frac{1}{e}\right)^n = \frac{1}{e} + \left(\frac{1}{e}\right)^2 + \left(\frac{1}{e}\right)^3 + \cdots$$

is a geometric series with initial term $a = 1/e$ and ratio $r = 1/e$, which converges to

$$\frac{1/e}{1-(1/e)} = \frac{1/e}{(e-1)/e} = \frac{1}{e-1}.$$

However, the related integral $\int_1^{\infty} (1/e)^x\,dx$ satisfies

$$\int_1^{\infty} \left(\frac{1}{e}\right)^x dx = \int_1^{\infty} e^{-x}\,dx = \lim_{b\to\infty} \int_1^b e^{-x}\,dx = \lim_{b\to\infty} -e^{-x}\Big|_1^b = \lim_{b\to\infty} \left[-e^{-b} + e^{-1}\right] = \frac{1}{e}.$$

Example 5.3.2: Using the Integral Test

For each of the following series, use the integral test to determine whether the series converges or diverges.

a. $\sum_{n=1}^{\infty} \frac{1}{n^3}$

b. $\sum_{n=1}^{\infty} \frac{1}{\sqrt{2n-1}}$

Solution

a. Compare

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \text{ and } \int_1^{\infty} \frac{1}{x^3}\,dx.$$

We have

$$\int_1^{\infty} \frac{1}{x^3}\,dx = \lim_{b\to\infty} \int_1^b \frac{1}{x^3}\,dx = \lim_{b\to\infty} \left[-\frac{1}{2x^2}\Big|_1^b\right] = \lim_{b\to\infty} \left[-\frac{1}{2b^2} + \frac{1}{2}\right] = \frac{1}{2}.$$

Thus the integral $\int_1^{\infty} \frac{1}{x^3}\,dx$ converges, and therefore so does the series

$$\sum_{n=1}^{\infty} \frac{1}{n^3}.$$

b. Compare

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{2n-1}} \text{ and } \int_1^{\infty} \frac{1}{\sqrt{2x-1}}\,dx.$$

Since

$$\int_1^{\infty} \frac{1}{\sqrt{2x-1}}\,dx = \lim_{b\to\infty} \int_1^{b} \frac{1}{\sqrt{2x-1}}\,dx = \lim_{b\to\infty} \sqrt{2x-1}\Big|_1^{b} = \lim_{b\to\infty} \left[\sqrt{2b-1}-1\right] = \infty,$$

the integral $\displaystyle\int_1^{\infty} \frac{1}{\sqrt{2x-1}}\,dx$ diverges, and therefore

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{2n-1}}$$

diverges.

? Exercise 5.3.2

Use the integral test to determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{n}{3n^2+1}$ converges or diverges.

Hint

Compare to the integral $\displaystyle\int_1^{\infty} \frac{x}{3x^2+1}\,dx.$

Answer

The series diverges.

The p-Series

The harmonic series $\displaystyle\sum_{n=1}^{\infty} 1/n$ and the series $\displaystyle\sum_{n=1}^{\infty} 1/n^2$ are both examples of a type of series called a p-series.

Definition: p-series

For any real number p, the series

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

is called a *p-series*.

We know the p-series converges if $p=2$ and diverges if $p=1$. What about other values of p? In general, it is difficult, if not impossible, to compute the exact value of most p-series. However, we can use the tests presented thus far to prove whether a p-series converges or diverges.

If $p<0$, then $1/n^p \to \infty$, and if $p=0$, then $1/n^p \to 1$. Therefore, by the divergence test,

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

diverges if $p \le 0$.

If $p>0$, then $f(x)=1/x^p$ is a positive, continuous, decreasing function. Therefore, for $p>0$, we use the integral test, comparing

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

and

$$\int_1^\infty \frac{1}{x^p}\,dx.$$

We have already considered the case when $p=1$. Here we consider the case when $p>0, p\neq 1$. For this case,

$$\int_1^\infty \frac{1}{x^p}\,dx=\lim_{b\to\infty}\int_1^b \frac{1}{x^p}\,dx=\lim_{b\to\infty}\frac{1}{1-p}x^{1-p}\,|_1^b=\lim_{b\to\infty}\frac{1}{1-p}[b^{1-p}-1].$$

Because

$$b^{1-p}\to 0 \text{ if } p>1 \text{ and } b^{1-p}\to\infty \text{ if } p<1,$$

we conclude that

$$\int_1^\infty \frac{1}{x^p}\,dx=\begin{cases}\dfrac{1}{p-1}, \text{if } p>1\\ \infty, \text{if } p<1.\end{cases}$$

Therefore, $\displaystyle\sum_{n=1}^\infty 1/n^p$ converges if $p>1$ and diverges if $0<p<1$.

In summary,

$$\sum_{n=1}^\infty \frac{1}{n^p}\begin{cases}\text{converges if } p>1\\ \text{diverges if } p\leq 1\end{cases}$$

.

Example 5.3.3: Testing for Convergence of p-series

For each of the following series, determine whether it converges or diverges.

a. $\displaystyle\sum_{n=1}^\infty \frac{1}{n^4}$

b. $\displaystyle\sum_{n=1}^\infty \frac{1}{n^{2/3}}$

Solution

a. This is a p-series with $p=4>1$, so the series converges.

b. Since $p=2/3<1$, the series diverges.

Exercise 5.3.3

Does the series $\displaystyle\sum_{n=1}^\infty \frac{1}{n^{5/4}}$ converge or diverge?

Hint

$p=5/4$

Answer

The series converges.

Estimating the Value of a Series

Suppose we know that a series $\displaystyle\sum_{n=1}^\infty a_n$ converges and we want to estimate the sum of that series. Certainly we can approximate that sum using any finite sum $\displaystyle\sum_{n=1}^N a_n$ where N is any positive integer. The question we address here is, for a convergent series $\displaystyle\sum_{n=1}^\infty a_n$,

how good is the approximation $\sum_{n=1}^{N} a_n$?

More specifically, if we let

$$R_N = \sum_{n=1}^{\infty} a_n - \sum_{n=1}^{N} a_n$$

be the remainder when the sum of an infinite series is approximated by the Nth partial sum, how large is R_N? For some types of series, we are able to use the ideas from the integral test to estimate R_N.

Note 5.3.1: Remainder Estimate from the Integral Test

Suppose $\sum_{n=1}^{\infty} a_n$ is a convergent series with positive terms. Suppose there exists a function f satisfying the following three conditions:

i. f is continuous,
ii. f is decreasing, and
iii. $f(n) = a_n$ for all integers $n \geq 1$.

Let S_N be the Nth partial sum of $\sum_{n=1}^{\infty} a_n$. For all positive integers N,

$$S_N + \int_{N+1}^{\infty} f(x)\,dx < \sum_{n=1}^{\infty} a_n < S_N + \int_{N}^{\infty} f(x)\,dx.$$

In other words, the remainder $R_N = \sum_{n=1}^{\infty} a_n - S_N = \sum_{n=N+1}^{\infty} a_n$ satisfies the following estimate:

$$\int_{N+1}^{\infty} f(x)\,dx < R_N < \int_{N}^{\infty} f(x)\,dx.$$

This is known as the **remainder estimate.**

We illustrate Note 5.3.1 in Figure 5.3.4. In particular, by representing the remainder $R_N = a_{N+1} + a_{N+2} + a_{N+3} + \cdots$ as the sum of areas of rectangles, we see that the area of those rectangles is bounded above by $\int_{N}^{\infty} f(x)\,dx$ and bounded below by $\int_{N+1}^{\infty} f(x)\,dx$. In other words,

$$R_N = a_{N+1} + a_{N+2} + a_{N+3} + \cdots > \int_{N+1}^{\infty} f(x)\,dx$$

and

$$R_N = a_{N+1} + a_{N+2} + a_{N+3} + \cdots < \int_{N}^{\infty} f(x)\,dx.$$

We conclude that

$$\int_{N+1}^{\infty} f(x)\,dx < R_N < \int_{N}^{\infty} f(x)\,dx.$$

Since

$$\sum_{n=1}^{\infty} a_n = S_N + R_N,$$

where S_N is the Nth partial sum, we conclude that

$$S_N + \int_{N+1}^{\infty} f(x)\,dx < \sum_{n=1}^{\infty} a_n < S_N + \int_{N}^{\infty} f(x)\,dx.$$

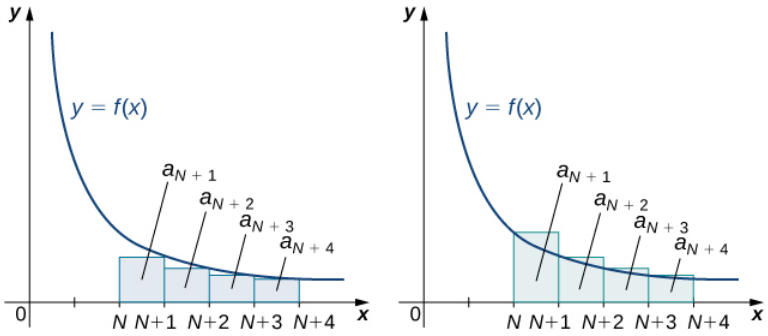


Figure 5.3.4: Given a continuous, positive, decreasing function f and a sequence of positive terms a_n such that $a_n = f(n)$ for all positive integers n, (a) the areas $a_{N+1} + a_{N+2} + a_{N+3} + \cdots < \int_N^{\infty} f(x)\,dx$, or (b) the areas $a_{N+1} + a_{N+2} + a_{N+3} + \cdots > \int_{N+1}^{\infty} f(x)\,dx$. Therefore, the integral is either an overestimate or an underestimate of the error.

Example 5.3.4: Estimating the Value of a Series

Consider the series $\sum_{n=1}^{\infty} \frac{1}{n^3}$.

a. Calculate $S_{10} = \sum_{n=1}^{10} \frac{1}{n^3}$ and estimate the error.

b. Determine the least value of N necessary such that S_N will estimate $\sum_{n=1}^{\infty} \frac{1}{n^3}$ to within 0.001.

Solution

a. Using a calculating utility, we have

$$S_{10} = 1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \cdots + \frac{1}{10^3} \approx 1.19753.$$

By the remainder estimate, we know

$$R_N < \int_N^{\infty} \frac{1}{x^3}\,dx.$$

We have

$$\int_{10}^{\infty} \frac{1}{x^3}\,dx = \lim_{b\to\infty} \int_{10}^{b} \frac{1}{x^3}\,dx = \lim_{b\to\infty} \left[-\frac{1}{2x^2}\right]_N^b = \lim_{b\to\infty} \left[-\frac{1}{2b^2} + \frac{1}{2N^2}\right] = \frac{1}{2N^2}.$$

Therefore, the error is $R_{10} < 1/2(10)^2 = 0.005$.

b. Find N such that $R_N < 0.001$. In part a. we showed that $R_N < 1/2N^2$. Therefore, the remainder $R_N < 0.001$ as long as $1/2N^2 < 0.001$. That is, we need $2N^2 > 1000$. Solving this inequality for N, we see that we need $N > 22.36$. To ensure that the remainder is within the desired amount, we need to round up to the nearest integer. Therefore, the minimum necessary value is $N = 23$.

Exercise 5.3.4

For $\sum_{n=1}^{\infty} \frac{1}{n^4}$, calculate S_5 and estimate the error R_5.

Hint

Use the remainder estimate $R_N < \int_N^{\infty} \frac{1}{x^4}\,dx$.

Answer

$S_5 \approx 1.09035, R_5 < 0.00267$

Key Concepts

- If $\lim_{n\to\infty} a_n \neq 0,$ then the series $\sum_{n=1}^{\infty} a_n$ diverges.
- If $\lim_{n\to\infty} a_n = 0,$ the series $\sum_{n=1}^{\infty} a_n$ may converge or diverge.
- If $\sum_{n=1}^{\infty} a_n$ is a series with positive terms a_n and f is a continuous, decreasing function such that $f(n) = a_n$ for all positive integers n, then

$$\sum_{n=1}^{\infty} a_n$$

and

$$\int_1^{\infty} f(x)\,dx$$

either both converge or both diverge. Furthermore, if $\sum_{n=1}^{\infty} a_n$ converges, then the Nth partial sum approximation S_N is accurate up to an error R_N where $\int_{N+1}^{\infty} f(x)\,dx < R_N < \int_N^{\infty} f(x)\,dx$.

- The p-series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \leq 1$.

Key Equations

- **Divergence test**

If $a_n \nrightarrow 0$ as $n \to \infty$, $\sum_{n=1}^{\infty} a_n$ diverges.

- ***p*-series**

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \quad \begin{cases} \text{converges, if } p > 1 \\ \text{diverges, if } p \leq 1 \end{cases}$$

- **Remainder estimate from the integral test**

$$\int_{N+1}^{\infty} f(x)\,dx < R_N < \int_N^{\infty} f(x)\,dx$$

Glossary

divergence test

if $\lim_{n\to\infty} a_n \neq 0,$ then the series $\sum_{n=1}^{\infty} a_n$ diverges

integral test

for a series $\sum_{n=1}^{\infty} a_n$ with positive terms a_n, if there exists a continuous, decreasing function f such that $f(n) = a_n$ for all positive integers n, then

$$\sum_{n=1}^{\infty} a_n$$

and

$$\int_1^{\infty} f(x)\,dx$$

either both converge or both diverge

_p_-series

a series of the form $\sum_{n=1}^{\infty} 1/n^p$

remainder estimate

for a series $\sum_{n=}^{\infty} 1a_n$ with positive terms a_n and a continuous, decreasing function f such that $f(n) = a_n$ for all positive integers n, the remainder $R_N = \sum_{n=1}^{\infty} a_n - \sum_{n=1}^{N} a_n$ satisfies the following estimate:

$$\int_{N+1}^{\infty} f(x)\,dx < R_N < \int_N^{\infty} f(x)\,dx$$

This page titled 5.3: The Divergence and Integral Tests is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.3: The Divergence and Integral Tests** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

5.4: Comparison Tests

Learning Objectives

- Use the comparison test to test a series for convergence.
- Use the limit comparison test to determine convergence of a series.

We have seen that the integral test allows us to determine the convergence or divergence of a series by comparing it to a related improper integral. In this section, we show how to use comparison tests to determine the convergence or divergence of a series by comparing it to a series whose convergence or divergence is known. Typically these tests are used to determine convergence of series that are similar to geometric series or p-series.

Comparison Test

In the preceding two sections, we discussed two large classes of series: geometric series and p-series. We know exactly when these series converge and when they diverge. Here we show how to use the convergence or divergence of these series to prove convergence or divergence for other series, using a method called the **comparison test.**

For example, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2+1}.$$

This series looks similar to the convergent series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

Since the terms in each of the series are positive, the sequence of partial sums for each series is monotone increasing. Furthermore, since

$$0 < \frac{1}{n^2+1} < \frac{1}{n^2}$$

for all positive integers n, the kth partial sum S_k of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2+1}$ satisfies

$$S_k = \sum_{n=1}^{k} \frac{1}{n^2+1} < \sum_{n=1}^{k} \frac{1}{n^2} < \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

(See Figure $5.4.1a$ and Table $5.4.1$.) Since the series on the right converges, the sequence S_k is bounded above. We conclude that S_k is a monotone increasing sequence that is bounded above. Therefore, by the Monotone Convergence Theorem, S_k converges, and thus

$$\sum_{n=1}^{\infty} \frac{1}{n^2+1}$$

converges.

Similarly, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n-1/2}.$$

This series looks similar to the divergent series

$$\sum_{n=1}^{\infty} \frac{1}{n}.$$

The sequence of partial sums for each series is monotone increasing and

$$\frac{1}{n-1/2} > \frac{1}{n} > 0$$

for every positive integer n. Therefore, the kth partial sum S_k of

$$\sum_{n=1}^{\infty} \frac{1}{n-1/2}$$

satisfies

$$S_k = \sum_{n=1}^{k} \frac{1}{n-1/2} > \sum_{n=1}^{k} \frac{1}{n}.$$

(See Figure $5.4.1n$and Table $5.4.1$). Since the series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges to infinity, the sequence of partial sums $\sum_{n=1}^{k} \frac{1}{n}$ is unbounded. Consequently, S_k is an unbounded sequence, and therefore diverges. We conclude that

$$\sum_{n=1}^{\infty} \frac{1}{n-1/2}$$

diverges.

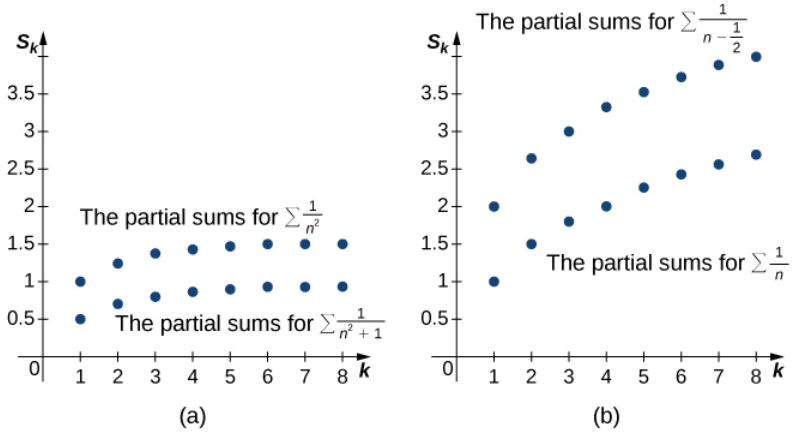


Figure 5.4.1: (a) Each of the partial sums for the given series is less than the corresponding partial sum for the converging $p-series$. (b) Each of the partial sums for the given series is greater than the corresponding partial sum for the diverging harmonic series.

Table 5.4.1: Comparing a series with a p-series ($p = 2$)

k	1	2	3	4	5	6	7	8
$\sum_{n=1}^{k} \frac{1}{n^2+1}$	0.5	0.7	0.8	0.8588	0.8973	0.9243	0.9443	0.9597
$\sum_{n=1}^{k} \frac{1}{n^2}$	1	1.25	1.3611	1.4236	1.4636	1.4914	1.5118	1.5274

Table 5.4.2: Comparing a series with the harmonic series

k	1	2	3	4	5	6	7	8
$\sum_{n=1}^{k} \frac{1}{n-1/2}$	2	2.6667	3.0667	3.3524	3.5746	3.7564	3.9103	4.0436

$\sum_{n=1}^{k} \frac{1}{n}$	1	1.5	1.8333	2.0933	2.2833	2.45	2.5929	2.7179

Comparison Test

i. Suppose there exists an integer N such that $0 \le a_n \le b_n$ for all $n \ge N$. If $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.

ii. Suppose there exists an integer N such that $a_n \ge b_n \ge 0$ for all $n \ge N$. If $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

Proof

We prove part i. The proof of part ii. is the contrapositive of part i. Let S_k be the sequence of partial sums associated with $\sum_{n=1}^{\infty} a_n$, and let $L = \sum_{n=1}^{\infty} b_n$. Since the terms $a_n \ge 0$,

$$S_k = a_1 + a_2 + \cdots + a_k \le a_1 + a_2 + \cdots + a_k + a_{k+1} = S_{k+1}.$$

Therefore, the sequence of partial sums is increasing. Further, since $a_n \le b_n$ for all $n \ge N$, then

$$\sum_{n=N}^{k} a_n \le \sum_{n=N}^{k} b_n \le \sum_{n=1}^{\infty} b_n = L.$$

Therefore, for all $k \ge 1$,

$$S_k = (a_1 + a_2 + \cdots + a_{N-1}) + \sum_{n=N}^{k} a_n \le (a_1 + a_2 + \cdots + a_{N-1}) + L.$$

Since $a_1 + a_2 + \cdots + a_{N-1}$ is a finite number, we conclude that the sequence S_k is bounded above. Therefore, S_k is an increasing sequence that is bounded above. By the Monotone Convergence Theorem, we conclude that S_k converges, and therefore the series $\sum_{n=1}^{\infty} a_n$ converges.

□

To use the comparison test to determine the convergence or divergence of a series $\sum_{n=1}^{\infty} a_n$, it is necessary to find a suitable series with which to compare it. Since we know the convergence properties of geometric series and p-series, these series are often used. If there exists an integer N such that for all $n \ge N$, each term an is less than each corresponding term of a known convergent series, then $\sum_{n=1}^{\infty} a_n$ converges. Similarly, if there exists an integer N such that for all $n \ge N$, each term an is greater than each corresponding term of a known divergent series, then $\sum_{n=1}^{\infty} a_n$ diverges.

Example 5.4.1: Using the Comparison Test

For each of the following series, use the comparison test to determine whether the series converges or diverges.

a. $\sum_{n=1}^{\infty} \frac{1}{n^3 + 3n + 1}$

b. $\sum_{n=1}^{\infty} \frac{1}{2^n + 1}$

c. $\sum_{n=2}^{\infty} \frac{1}{\ln n}$

Solution

a. Compare to $\sum_{n=1}^{\infty} \frac{1}{n^3}$. Since $\sum_{n=1}^{\infty} \frac{1}{n^3}$ is a p-series with $p = 3$, it converges. Further,

$$\frac{1}{n^3+3n+1} < \frac{1}{n^3}$$

for every positive integer n. Therefore, we can conclude that $\sum_{n=1}^{\infty} \frac{1}{n^3+3n+1}$ converges.

b. Compare to $\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n$. Since $\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n$ is a geometric series with $r = \frac{1}{2}$ and $\left|\frac{1}{2}\right| < 1$, it converges. Also,

$$\frac{1}{2^n+1} < \frac{1}{2^n}$$

for every positive integer n. Therefore, we see that $\sum_{n=1}^{\infty} \frac{1}{2^n+1}$ converges.

c. Compare to $\sum_{n=2}^{\infty} \frac{1}{n}$. Since

$$\frac{1}{\ln n} > \frac{1}{n}$$

for every integer $n \geq 2$ and $\sum_{n=2}^{\infty} \frac{1}{n}$ diverges, we have that $\sum_{n=2}^{\infty} \frac{1}{\ln n}$ diverges.

? Exercise 5.4.1

Use the comparison test to determine if the series $\sum_{n=1}^{\infty} \frac{n}{n^3+n+1}$ converges or diverges.

Hint

Find a value p such that $\frac{n}{n^3+n+1} \leq \frac{1}{n^p}$.

Answer

The series converges.

Limit Comparison Test

The comparison test works nicely if we can find a comparable series satisfying the hypothesis of the test. However, sometimes finding an appropriate series can be difficult. Consider the series

$$\sum_{n=2}^{\infty} \frac{1}{n^2-1}.$$

It is natural to compare this series with the convergent series

$$\sum_{n=2}^{\infty} \frac{1}{n^2}.$$

However, this series does not satisfy the hypothesis necessary to use the comparison test because

$$\frac{1}{n^2-1} > \frac{1}{n^2}$$

for all integers $n \geq 2$. Although we could look for a different series with which to compare $\sum_{n=2}^{\infty} \frac{1}{n^2-1}$, instead we show how we can use the **limit comparison test** to compare

$$\sum_{n=2}^{\infty} \frac{1}{n^2-1}$$

and

$$\sum_{n=2}^{\infty} \frac{1}{n^2}.$$

Let us examine the idea behind the limit comparison test. Consider two series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$. with positive terms a_n and b_n and evaluate

$$\lim_{n\to\infty} \frac{a_n}{b_n}.$$

If

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L \neq 0,$$

then, for n sufficiently large, $a_n \approx Lb_n$. Therefore, either both series converge or both series diverge. For the series $\sum_{n=2}^{\infty} \frac{1}{n^2-1}$ and $\sum_{n=2}^{\infty} \frac{1}{n^2}$, we see that

$$\lim_{n\to\infty} \frac{1/(n^2-1)}{1/n^2} = \lim_{n\to\infty} \frac{n^2}{n^2-1} = 1.$$

Since $\sum_{n=2}^{\infty} \frac{1}{n^2}$ converges, we conclude that

$$\sum_{n=2}^{\infty} \frac{1}{n^2-1}$$

converges.

The limit comparison test can be used in two other cases. Suppose

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

In this case, a_n/b_n is a bounded sequence. As a result, there exists a constant M such that $a_n \leq Mb_n$. Therefore, if $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges. On the other hand, suppose

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \infty.$$

In this case, a_n/b_n is an unbounded sequence. Therefore, for every constant M there exists an integer N such that $a_n \geq Mb_n$ for all $n \geq N$. Therefore, if $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges as well.

Limit Comparison Test

Let $a_n, b_n \geq 0$ for all $n \geq 1$.

i. If $\lim\limits_{n\to\infty} \dfrac{a_n}{b_n} = L \neq 0,$ then $\sum\limits_{n=1}^{\infty} a_n$ and $\sum\limits_{n=1}^{\infty} b_n$ both converge or both diverge.

ii. If $\lim\limits_{n\to\infty} \dfrac{a_n}{b_n} = 0$ and $\sum\limits_{n=1}^{\infty} b_n$ converges, then $\sum\limits_{n=1}^{\infty} a_n$ converges.

iii. If $\lim\limits_{n\to\infty} \dfrac{a_n}{b_n} = \infty$ and $\sum\limits_{n=1}^{\infty} b_n$ diverges, then $\sum\limits_{n=1}^{\infty} a_n$ diverges.

Note that if $\dfrac{a_n}{b_n} \to 0$ and $\sum\limits_{n=1}^{\infty} b_n$ diverges, the limit comparison test gives no information. Similarly, if $\dfrac{a_n}{b_n} \to \infty$ and $\sum\limits_{n=1}^{\infty} b_n$ converges, the test also provides no information. For example, consider the two series $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$ and $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^2}$. These series are both p-series with $p = \frac{1}{2}$ and $p = 2$, respectively. Since $p = \frac{1}{2} < 1,$ the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$ diverges. On the other hand, since $p = 2 > 1,$ the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^2}$ converges. However, suppose we attempted to apply the limit comparison test, using the convergent $\boldsymbol{p}$**−series** $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^3}$ as our comparison series. First, we see that

$$\frac{1/\sqrt{n}}{1/n^3} = \frac{n^3}{\sqrt{n}} = n^{5/2} \to \infty \text{ as } n \to \infty.$$

Similarly, we see that

$$\frac{1/n^2}{1/n^3} = n \to \infty \text{ as } n \to \infty.$$

Therefore, if $\dfrac{a_n}{b_n} \to \infty$ when $\sum\limits_{n=1}^{\infty} b_n$ converges, we do not gain any information on the convergence or divergence of $\sum\limits_{n=1}^{\infty} a_n$.

Example 5.4.2: Using the Limit Comparison Test

For each of the following series, use the limit comparison test to determine whether the series converges or diverges. If the test does not apply, say so.

a. $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}+1}$

b. $\sum\limits_{n=1}^{\infty} \dfrac{2^n+1}{3^n}$

c. $\sum\limits_{n=1}^{\infty} \dfrac{\ln(n)}{n^2}$

Solution

a. Compare this series to $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$. Calculate

$$\lim_{n\to\infty} \frac{1/(\sqrt{n}+1)}{1/\sqrt{n}} = \lim_{n\to\infty} \frac{\sqrt{n}}{\sqrt{n}+1} = \lim_{n\to\infty} \frac{1/\sqrt{n}}{1+1/\sqrt{n}} = 1.$$

By the limit comparison test, since $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$ diverges, then $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}+1}$ diverges.

b. Compare this series to $\sum_{n=1}^{\infty}\left(\frac{2}{3}\right)^n$. We see that

$$\lim_{n\to\infty}\frac{(2^n+1)/3^n}{2^n/3^n}=\lim_{n\to\infty}\frac{2^n+1}{3^n}\cdot\frac{3^n}{2^n}=\lim_{n\to\infty}\frac{2^n+1}{2^n}=\lim_{n\to\infty}\left[1+\left(\tfrac{1}{2}\right)^n\right]=1.$$

Therefore,

$$\lim_{n\to\infty}\frac{(2^n+1)/3^n}{2^n/3^n}=1.$$

Since $\sum_{n=1}^{\infty}\left(\frac{2}{3}\right)^n$ converges, we conclude that $\sum_{n=1}^{\infty}\frac{2^n+1}{3^n}$ converges.

c. Since $\ln n<n,$ compare with $\sum_{n=1}^{\infty}\frac{1}{n}$. We see that

$$\lim_{n\to\infty}\frac{\ln n/n^2}{1/n}=\lim_{n\to\infty}\frac{\ln n}{n^2}\cdot\frac{n}{1}=\lim_{n\to\infty}\frac{\ln n}{n}.$$

In order to evaluate $\lim_{n\to\infty}\ln n/n$, evaluate the limit as $x\to\infty$ of the real-valued function $\ln(x)/x$. These two limits are equal, and making this change allows us to use L'Hôpital's rule. We obtain

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

Therefore, $\lim_{n\to\infty}\frac{\ln n}{n}=0$, and, consequently,

$$\lim_{n\to\infty}\frac{(\ln n)/n^2}{1/n}=0.$$

Since the limit is 0 but $\sum_{n=1}^{\infty}\frac{1}{n}$ diverges, the limit comparison test does not provide any information.

Compare with $\sum_{n=1}^{\infty}\frac{1}{n^2}$ instead. In this case,

$$\lim_{n\to\infty}\frac{(\ln n)/n^2}{1/n^2}=\lim_{n\to\infty}\frac{\ln n}{n^2}\cdot\frac{n^2}{1}=\lim_{n\to\infty}\ln n=\infty.$$

Since the limit is ∞ but $\sum_{n=1}^{\infty}\frac{1}{n^2}$ converges, the test still does not provide any information.

So now we try a series between the two we already tried. Choosing the series $\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}$, we see that

$$\lim_{n\to\infty}\frac{(\ln n)/n^2}{1/n^{3/2}}=\lim_{n\to\infty}\frac{\ln n}{n^2}\cdot\frac{n^{3/2}}{1}=\lim_{n\to\infty}\frac{\ln n}{\sqrt{n}}\ .$$

As above, in order to evaluate $\lim_{n\to\infty}\frac{\ln n}{\sqrt{n}}$, evaluate the limit as $x\to\infty$ of the real-valued function $\frac{\ln n}{\sqrt{n}}$. Using L'Hôpital's rule,

$$\lim_{x\to\infty}\frac{\ln x}{\sqrt{x}}=\lim_{x\to\infty}\frac{2\sqrt{x}}{x}=\lim_{x\to\infty}\frac{2}{\sqrt{x}}=0\ .$$

Since the limit is 0 and $\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}$ converges, we can conclude that $\sum_{n=1}^{\infty}\frac{\ln n}{n^2}$ converges.

Exercise 5.4.2

Use the limit comparison test to determine whether the series $\sum_{n=1}^{\infty} \frac{5^n}{3^n+2}$ converges or diverges.

Hint

Compare with a geometric series.

Answer

The series diverges.

Key Concepts

- The comparison tests are used to determine convergence or divergence of series with positive terms.
- When using the comparison tests, a series $\sum_{n=1}^{\infty} a_n$ is often compared to a geometric or p-series.

Glossary

comparison test

If $0 \le a_n \le b_n$ for all $n \ge N$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges; if $a_n \ge b_n \ge 0$ for all $n \ge N$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

limit comparison test

Suppose $a_n, b_n \ge 0$ for all $n \ge 1$. If $\lim_{n\to\infty} a_n/b_n \to L \ne 0$, then $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge or both diverge; if $\lim_{n\to\infty} a_n/b_n \to 0$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges. If $\lim_{n\to\infty} a_n/b_n \to \infty$, and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.

This page titled 5.4: Comparison Tests is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.4: Comparison Tests** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

5.5: Alternating Series

Learning Objectives

- Use the alternating series test to test an alternating series for convergence.
- Estimate the sum of an alternating series.
- Explain the meaning of absolute convergence and conditional convergence.

So far in this chapter, we have primarily discussed series with positive terms. In this section we introduce alternating series—those series whose terms alternate in sign. We will show in a later chapter that these series often arise when studying power series. After defining alternating series, we introduce the alternating series test to determine whether such a series converges.

The Alternating Series Test

A series whose terms alternate between positive and negative values is an **alternating series.** For example, the series

$$\sum_{n=1}^{\infty}\left(-\frac{1}{2}\right)^n = -\frac{1}{2}+\frac{1}{4}-\frac{1}{8}+\frac{1}{16}-\ldots \tag{5.5.1}$$

and

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n} = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots \tag{5.5.2}$$

are both alternating series.

Definition: Alternating Series

Any series whose terms alternate between positive and negative values is called an *alternating series*. An alternating series can be written in the form

$$\sum_{n=1}^{\infty}(-1)^{n+1}b_n = b_1-b_2+b_3-b_4+\ldots \tag{5.5.3}$$

or

$$\sum_{n-1}^{\infty}(-1)^{n}b_n = -b_1+b_2-b_3+b_4-\ldots \tag{5.5.4}$$

Where $b_n \geq 0$ for all positive integers n.

Series (1), shown in Equation 5.5.1, is a *geometric series*. Since $|r| = |-1/2| < 1,$ the series converges. Series (2), shown in Equation 5.5.2, is called the *alternating harmonic series*. We will show that whereas the harmonic series diverges, the alternating harmonic series converges. To prove this, we look at the sequence of partial sums $\{S_k\}$ (Figure 5.5.1).

Proof

Consider the odd terms S_{2k+1} for $k \geq 0$. Since $1/(2k+1) < 1/2k$,

$$S_{2k+1} = S_{2k-1} - \frac{1}{2k}+\frac{1}{2k+1} < S_{2k-1}.$$

Therefore, $\{S_{2k+1}\}$ is a decreasing sequence. Also,

$$S_{2k+1} = \left(1-\frac{1}{2}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\ldots+\left(\frac{1}{2k-1}-\frac{1}{2k}\right)+\frac{1}{2k+1} > 0.$$

Therefore, $\{S_{2k+1}\}$ is bounded below. Since $\{S_{2k+1}\}$ is a decreasing sequence that is bounded below, by the Monotone Convergence Theorem, $\{S_{2k+1}\}$ converges. Similarly, the even terms $\{S_{2k}\}$ form an increasing sequence that is bounded

above because

$$S_{2k} = S_{2k-2} + \frac{1}{2k-1} - \frac{1}{2k} > S_{2k-2}$$

and

$$S_{2k} = 1 + \left(-\frac{1}{2} + \frac{1}{3}\right) + \ldots + \left(-\frac{1}{2k-2} + \frac{1}{2k-1}\right) - \frac{1}{2k} < 1.$$

Therefore, by the Monotone Convergence Theorem, the sequence $\{S_{2k}\}$ also converges. Since

$$S_{2k+1} = S_{2k} + \frac{1}{2k+1},$$

we know that

$$\lim_{k\to\infty} S_{2k+1} = \lim_{k\to\infty} S_{2k} + \lim_{k\to\infty} \frac{1}{2k+1}.$$

Letting $S = \lim_{k\to\infty} S_{2k+1}$ and using the fact that $1/(2k+1) \to 0$, we conclude that $\lim_{k\to\infty} S_{2k} = S$. Since the odd terms and the even terms in the sequence of partial sums converge to the same limit S, it can be shown that the sequence of partial sums converges to S, and therefore the alternating harmonic series converges to S.

It can also be shown that $S = \ln 2$, and we can write

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + a \ldots = \ln(2).$$

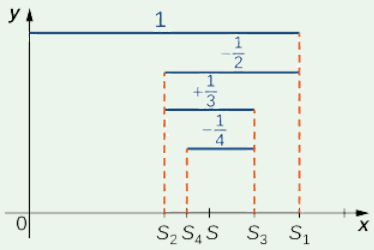


Figure 5.5.1: For the alternating harmonic series, the odd terms S_{2k+1} in the sequence of partial sums are decreasing and bounded below. The even terms S_{2k} are increasing and bounded above.

□

More generally, any alternating series of form (3) (Equation 5.5.3) or (4) (Equation 5.5.4) converges as long as $b_1 \geq b_2 \geq b_3 \geq \cdots$ and $b_n \to 0$ (Figure 5.5.2). The proof is similar to the proof for the alternating harmonic series.

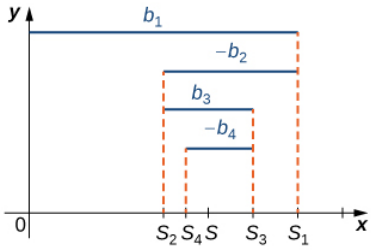


Figure 5.5.2: For an alternating series $b_1 - b_2 + b_3 - \cdots$ in which $b_1 > b_2 > b_3 > \cdots$, the odd terms S_{2k+1} in the sequence of partial sums are decreasing and bounded below. The even terms S_{2k} are increasing and bounded above.

Alternating Series Test

An alternating series of the form

$$\sum_{n=1}^{\infty} (-1)^{n+1} b_n$$

or

$$\sum_{n=1}^{\infty}(-1)^n b_n$$

converges if

i. $0 \le b_{n+1} \le b_n$ for all $n \ge 1$ and
ii. $\lim_{n\to\infty} b_n = 0$.

This is known as the *alternating series test*.

We remark that this theorem is true more generally as long as there exists some integer N such that $0 \le b_{n+1} \le b_n$ for all $n \ge N$.

Example 5.5.1: Convergence of Alternating Series

For each of the following alternating series, determine whether the series converges or diverges.

a. $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2}$

b. $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{n}{n+1}$

Solution

a. Since $\frac{1}{(n+1)^2} < \frac{1}{n^2}$ and $\frac{1}{n^2} \to 0$, the series converges.

b. Since $n/(n+1) \nrightarrow 0$ as $n \to \infty$, we cannot apply the alternating series test. Instead, we use the nth term test for divergence. Since $\lim_{n\to\infty} \frac{n}{n+1} = 1 \ne 0$, the series diverges.

Exercise 5.5.1

Determine whether the series $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{n}{2^n}$ converges or diverges.

Hint

Is $\left\{\frac{n}{2^n}\right\}$ decreasing? What is $\lim_{n\to\infty}\frac{n}{2^n}$?

Answer

The series converges.

Remainder of an Alternating Series

It is difficult to explicitly calculate the sum of most alternating series, so typically the sum is approximated by using a partial sum. When doing so, we are interested in the amount of error in our approximation. Consider an alternating series

$$\sum_{n=1}^{\infty}(-1)^{n+1} b_n$$

satisfying the hypotheses of the alternating series test. Let S denote the sum of this series and S_k be the corresponding sequence of partial sums. From Figure 5.5.2, we see that for any integer $N \ge 1$, the remainder R_N satisfies

$$|R_N| = |S - S_N| \le |S_{N+1} - S_N| = b_{n+1}.$$

Remainders in Alternating Series

Consider an alternating series of the form

$$\sum_{n=1}^{\infty}(-1)^{n+1}b_n$$

or

$$\sum_{n=1}^{\infty}(-1)^{n}b_n$$

that satisfies the hypotheses of the alternating series test. Let S denote the sum of the series and S_N denote the Nth partial sum. For any integer $N \geq 1$, the remainder $R_N = S - S_N$ satisfies

$$|R_N| \leq b_{N+1}.$$

In other words, if the conditions of the alternating series test apply, then the error in approximating the infinite series by the Nth partial sum S_N is in magnitude at most the size of the next term b_{N+1}.

Example 5.5.2: Estimating the Remainder of an Alternating Series

Consider the alternating series

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}.$$

Use the remainder estimate to determine a bound on the error R_{10} if we approximate the sum of the series by the partial sum S_{10}.

Solution

From the theorem stated above,

$$|R_{10}| \leq b_{11} = \frac{1}{11^2} \approx 0.008265.$$

Exercise 5.5.2

Find a bound for R_{20} when approximating $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}$ by S_{20}.

Hint

$|R_{20}| \leq b_{21}$

Answer

0.04762

Absolute and Conditional Convergence

Consider a series $\sum_{n=1}^{\infty}a_n$ and the related series $\sum_{n=1}^{\infty}|a_n|$. Here we discuss possibilities for the relationship between the convergence of these two series. For example, consider the alternating harmonic series $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}$. The series whose terms are the absolute

value of these terms is the harmonic series, since $\sum_{n=1}^{\infty}\left|\frac{(-1)^{n+1}}{n}\right|=\sum_{n=1}^{\infty}\frac{1}{n}$. Since the alternating harmonic series converges, but the harmonic series diverges, we say the alternating harmonic series exhibits conditional convergence.

By comparison, consider the series $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}$. The series whose terms are the absolute values of the terms of this series is the series $\sum_{n=1}^{\infty}\frac{1}{n^2}$. Since both of these series converge, we say the series $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}$ exhibits absolute convergence.

Definition: Absolute Convergence

A series $\sum_{n=1}^{\infty}a_n$ exhibits *absolute convergence* if $\sum_{n=1}^{\infty}|a_n|$ converges. A series $\sum_{n=1}^{\infty}a_n$ exhibits *conditional convergence* if $\sum_{n=1}^{\infty}a_n$ converges but $\sum_{n=1}^{\infty}|a_n|$ diverges.

As shown by the alternating harmonic series, a series $\sum_{n=1}^{\infty}a_n$ may converge, but $\sum_{n=1}^{\infty}|a_n|$ may diverge. In the following theorem, however, we show that if $\sum_{n=1}^{\infty}|a_n|$ converges, then $\sum_{n=1}^{\infty}a_n$ converges.

Absolute Convergence Implies Convergence

If $\sum_{n=1}^{\infty}|a_n|$ converges, then $\sum_{n=1}^{\infty}a_n$ converges.

Proof

Suppose that $\sum_{n=1}^{\infty}|a_n|$ converges. We show this by using the fact that $a_n=|a_n|$ or $a_n=-|a_n|$ and therefore $|a_n|+a_n=2|a_n|$ or $|a_n|+a_n=0$. Therefore, $0\le|a_n|+a_n\le 2|a_n|$. Consequently, by the comparison test, since $2\sum_{n=1}^{\infty}|a_n|$ converges, the series

$$\sum_{n=1}^{\infty}(|a_n|+a_n)$$

converges. By using the algebraic properties for convergent series, we conclude that

$$\sum_{n=1}^{\infty}a_n=\sum_{n=1}^{\infty}(|a_n|+a_n)-\sum_{n=1}^{\infty}|a_n|$$

converges.

□

Example 5.5.3: Absolute versus Conditional Convergence

For each of the following series, determine whether the series converges absolutely, converges conditionally, or diverges.

a. $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{3n+1}$

b. $\sum_{n=1}^{\infty}\frac{\cos(n)}{n^2}$

Solution

a. We can see that

$$\sum_{n=1}^{\infty}\left|\frac{(-1)^{n+1}}{3n+1}\right|=\sum_{n=1}^{\infty}\frac{1}{3n+1}$$

diverges by using the limit comparison test with the harmonic series. In fact,

$$\lim_{n\to\infty}\frac{1/(3n+1)}{1/n}=\frac{1}{3}.$$

Therefore, the series does not converge absolutely. However, since

$$\frac{1}{3(n+1)+1}<\frac{1}{3n+1}\text{ and }\frac{1}{3n+1}\to 0,$$

the series converges. We can conclude that $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{3n+1}$ converges conditionally.

b. Noting that $|\cos n|\le 1$, to determine whether the series converges absolutely, compare

$$\sum_{n=1}^{\infty}\left|\frac{\cos n}{n^2}\right|$$

with the series $\sum_{n=1}^{\infty}\frac{1}{n^2}$. Since $\sum_{n=1}^{\infty}\frac{1}{n^2}$ converges, by the comparison test, $\sum_{n=1}^{\infty}\left|\frac{\cos n}{n^2}\right|$ converges, and therefore $\sum_{n=1}^{\infty}\frac{\cos n}{n^2}$ converges absolutely.

? Exercise 5.5.3

Determine whether the series $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{n}{2n^3+1}$ converges absolutely, converges conditionally, or diverges.

Hint

Check for absolute convergence first.

Answer

The series converges absolutely.

To see the difference between absolute and conditional convergence, look at what happens when we rearrange the terms of the alternating harmonic series $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}$. We show that we can rearrange the terms so that the new series diverges. Certainly if we rearrange the terms of a finite sum, the sum does not change. When we work with an infinite sum, however, interesting things can happen.

Begin by adding enough of the positive terms to produce a sum that is larger than some real number $M=10$ For example, let $M=10$, and find an integer k such that

$$1+\frac{1}{3}+\frac{1}{5}+\cdots+\frac{1}{2k-1}>10$$

(We can do this because the series $\sum_{n=1}^{\infty}\frac{1}{2n-1}$ diverges to infinity.) Then subtract $1/2$. Then add more positive terms until the sum reaches 100. That is, find another integer $j>k$ such that

$$(1+\frac{1}{3}+\cdots+\frac{1}{2k-1}-\frac{1}{2}+\frac{1}{2k+1}+\ldots+\frac{1}{2j+1}>100.$$

Then subtract $1/4$. Continuing in this way, we have found a way of rearranging the terms in the alternating harmonic series so that the sequence of partial sums for the rearranged series is unbounded and therefore diverges.

The terms in the alternating harmonic series can also be rearranged so that the new series converges to a different value. In Example, we show how to rearrange the terms to create a new series that converges to $3\ln(2)/2$. We point out that the alternating harmonic series can be rearranged to create a series that converges to any real number r; however, the proof of that fact is beyond the scope of this text.

In general, any series $\sum_{n=1}^{\infty} a_n$ that converges conditionally can be rearranged so that the new series diverges or converges to a different real number. A series that converges absolutely does not have this property. For any series $\sum_{n=1}^{\infty} a_n$ that converges absolutely, the value of $\sum_{n=1}^{\infty} a_n$ is the same for any rearrangement of the terms. This result is known as the **Riemann Rearrangement Theorem**, which is beyond the scope of this book.

Example 5.5.4: Rearranging Series

Use the fact that

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\cdots=\ln 2$$

to rearrange the terms in the alternating harmonic series so the sum of the rearranged series is $3\ln(2)/2$.

Solution

Let

$$\sum_{n=1}^{\infty} a_n = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\cdots.$$

Since $\sum_{n=1}^{\infty} a_n = \ln(2)$, by the algebraic properties of convergent series,

$$\sum_{n=1}^{\infty} \frac{1}{2}a_n = \frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\cdots=\frac{1}{2}\sum_{n=1}^{\infty} a_n = \frac{\ln 2}{2}.$$

Now introduce the series $\sum_{n=1}^{\infty} b_n$ such that for all $n \geq 1, b_{2n-1}=0$ and $b_{2n}=a_n/2$. Then

$$\sum_{n=1}^{\infty} b_n = 0+\frac{1}{2}+0-\frac{1}{4}+0+\frac{1}{6}+0-\frac{1}{8}+\cdots=\frac{\ln 2}{2}.$$

Then using the algebraic limit properties of convergent series, since $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge, the series $\sum_{n=1}^{\infty}(a_n+b_n)$ converges and

$$\sum_{n=1}^{\infty}(a_n+b_n)=\sum_{n=1}^{\infty} a_n+\sum_{n=1}^{\infty} b_n=\ln 2+\frac{\ln 2}{2}=\frac{3\ln 2}{2}.$$

Now adding the corresponding terms, a_n and b_n, we see that

$$\sum_{n=1}^{\infty}(a_n+b_n)=(1+0)+\left(-\frac{1}{2}+\frac{1}{2}\right)+\left(\frac{1}{3}+0\right)+\left(-\frac{1}{4}-14\right)+\left(\frac{1}{5}+0\right)+\left(-\frac{1}{6}+\frac{1}{6}\right)+\left(\frac{1}{7}+0\right)$$
$$+\left(\frac{1}{8}-\frac{1}{8}\right)+\cdots=1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\cdots.$$

We notice that the series on the right side of the equal sign is a rearrangement of the alternating harmonic series. Since $\sum_{n=1}^{\infty}(a_n+b_n)=3\ln(2)/2,$ we conclude that

$$1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\cdots=\frac{3\ln(2)}{2}.$$

Therefore, we have found a rearrangement of the alternating harmonic series having the desired property.

Key Concepts

- For an alternating series $\sum_{n=1}^{\infty}(-1)^{n+1}b_n$, if $b_{k+1}\leq b_k$ for all k and $b_k\to 0$ as $k\to\infty$, the alternating series converges.
- If $\sum_{n=1}^{\infty}|a_n|$ converges, then $\sum_{n=1}^{\infty}a_n$ converges.

Key Equations

- **Alternating series**

$$\sum_{n=1}^{\infty}(-1)^{n+1}b_n=b_1-b_2+b_3-b_4+\cdots \quad \text{or}$$

$$\sum_{n=1}^{\infty}(-1)^{n}b_n=-b_1+b_2-b_3+b_4-\cdots$$

Glossary

absolute convergence

if the series $\sum_{n=1}^{\infty}|a_n|$ converges, the series $\sum_{n=1}^{\infty}a_n$ is said to converge absolutely

alternating series

a series of the form $\sum_{n=1}^{\infty}(-1)^{n+1}b_n$ or $\sum_{n=1}^{\infty}(-1)^{n}b_n$, where $b_n\geq 0$, is called an alternating series

alternating series test

for an alternating series of either form, if $b_{n+1}\leq b_n$ for all integers $n\geq 1$ and $b_n\to 0$, then an alternating series converges

conditional convergence

if the series $\sum_{n=1}^{\infty}a_n$ converges, but the series $\sum_{n=1}^{\infty}|a_n|$ diverges, the series $\sum_{n=1}^{\infty}a_n$ is said to converge conditionally

This page titled 5.5: Alternating Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.5: Alternating Series** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

5.6: Ratio and Root Tests

Learning Objectives

- Use the ratio test to determine absolute convergence of a series.
- Use the root test to determine absolute convergence of a series.
- Describe a strategy for testing the convergence of a given series.

In this section, we prove the last two series convergence tests: the ratio test and the root test. These tests are particularly nice because they do not require us to find a comparable series. The ratio test will be especially useful in the discussion of power series in the next chapter. Throughout this chapter, we have seen that no single convergence test works for all series. Therefore, at the end of this section we discuss a strategy for choosing which convergence test to use for a given series.

Ratio Test

Consider a series $\sum_{n=1}^{\infty} a_n$. From our earlier discussion and examples, we know that $\lim_{n\to\infty} a_n = 0$ is not a sufficient condition for the series to converge. Not only do we need $a_n \to 0$, but we need $a_n \to 0$ quickly enough. For example, consider the series $\sum_{n=1}^{\infty} \frac{1}{n}$ and the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$. We know that $\frac{1}{n} \to 0$ and $\frac{1}{n^2} \to 0$. However, only the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges. The series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges because the terms in the sequence $\left\{\frac{1}{n}\right\}$ do not approach zero fast enough as $n \to \infty$. Here we introduce the **ratio test,** which provides a way of measuring how fast the terms of a series approach zero.

Ratio Test

Let $\sum_{n=1}^{\infty} a_n$ be a series with nonzero terms. Let

$$\rho = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|.$$

i. If $0 \le \rho < 1$, then $\sum_{n=1}^{\infty} a_n$ converges absolutely.

ii. If $\rho > 1$ or $\rho = \infty$, then $\sum_{n=1}^{\infty} a_n$ diverges.

iii. If $\rho = 1$, the test does not provide any information.

Proof

Let $\sum_{n=1}^{\infty} a_n$ be a series with nonzero terms.

We begin with the proof of part i. In this case, $\rho = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$. Since $0 \le \rho < 1$, there exists R such that $0 \le \rho < R < 1$. Let $\varepsilon = R - \rho > 0$. By the definition of limit of a sequence, there exists some integer N such that

$$\left|\left|\frac{a_{n+1}}{a_n}\right| - \rho\right| < \varepsilon, \text{ for all } n \ge N.$$

Therefore,

$$\left|\frac{a_{n+1}}{a_n}\right| < \rho + \varepsilon = R, \text{ for all } n \ge N$$

and, thus,

$$
\begin{aligned}
&|a_{N+1}| < R|a_N| \\
&|a_{N+2}| < R|a_{N+1}| < R^2|a_N| \\
&|a_{N+3}| < R|a_{N+2}| < R^2|a_{N+1}| < R^3|a_N| \\
&|a_{N+4}| < R|a_{N+3}| < R^2|a_{N+2}| < R^3|a_{N+1}| < R^4|a_N| \\
&\vdots
\end{aligned}
$$

Since $R < 1$, the geometric series

$$R|a_N| + R^2|a_N| + R^3|a_N| + \cdots$$

converges. Given the inequalities above, we can apply the comparison test and conclude that the series

$$|a_{N+1}| + |a_{N+2}| + |a_{N+3}| + |a_{N+4}| + \cdots$$

converges. Therefore, since

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{N} |a_n| + \sum_{n=N+1}^{\infty} |a_n|$$

where $\sum_{n=1}^{N} |a_n|$ is a finite sum and $\sum_{n=N+1}^{\infty} |a_n|$ converges, we conclude that $\sum_{n=1}^{\infty} |a_n|$ converges.

For part ii.

$$\rho = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| > 1.$$

Since $\rho > 1$, there exists R such that $\rho > R > 1$. Let $\varepsilon = \rho - R > 0$. By the definition of the limit of a sequence, there exists an integer N such that

$$\left|\left|\frac{a_{n+1}}{a_n}\right| - \rho\right| < \varepsilon, \text{ for all } n \geq N.$$

Therefore,

$$R = \rho - \varepsilon < \left|\frac{a_{n+1}}{a_n}\right|, \text{ for all } n \geq N,$$

and, thus,

$$
\begin{aligned}
&|a_{N+1}| > R|a_N| \\
&|a_{N+2}| > R|a_{N+1}| > R^2|a_N| \\
&|a_{N+3}| > R|a_{N+2}| > R^2|a_{N+1}| > R^3|a_N| \\
&|a_{N+4}| > R|a_{N+3}| > R^2|a_{N+2}| > R^3|a_{N+1}| > R^4|a_N|.
\end{aligned}
$$

Since $R > 1$, the geometric series

$$R|a_N| + R^2|a_N| + R^3|a_N| + \cdots$$

diverges. Applying the comparison test, we conclude that the series

$$|a_{N+1}| + |a_{N+2}| + |a_{N+3}| + \cdots$$

diverges, and therefore the series $\sum_{n=1}^{\infty} |a_n|$ diverges.

For part iii. we show that the test does not provide any information if $\rho = 1$ by considering the $p-series$ $\sum_{n=1}^{\infty} \frac{1}{n^p}$. For any real number p,

$$\rho = \lim_{n\to\infty} \frac{1/(n+1)^p}{1/n^p} = \lim_{n\to\infty} \frac{n^p}{(n+1)^p} = 1.$$

However, we know that if $p \le 1$, the **p−series** $\sum_{n=1}^{\infty} \frac{1}{n^p}$ diverges, whereas $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$.

□

The ratio test is particularly useful for series whose terms contain factorials or exponential, where the ratio of terms simplifies the expression. The ratio test is convenient because it does not require us to find a comparative series. The drawback is that the test sometimes does not provide any information regarding convergence.

Example 5.6.1: Using the Ratio Test

For each of the following series, use the ratio test to determine whether the series converges or diverges.

a. $\sum_{n=1}^{\infty} \frac{2^n}{n!}$

b. $\sum_{n=1}^{\infty} \frac{n^n}{n!}$

c. $\sum_{n=1}^{\infty} \frac{(-1)^n (n!)^2}{(2n)!}$

Solution

a. From the ratio test, we can see that

$$\rho = \lim_{n\to\infty} \frac{2^{n+1}/(n+1)!}{2^n/n!} = \lim_{n\to\infty} \frac{2^{n+1}}{(n+1)!} \cdot \frac{n!}{2^n}.$$

Since $(n+1)! = (n+1)\cdot n!$,

$$\rho = \lim_{n\to\infty} \frac{2}{n+1} = 0.$$

Since $\rho < 1$, the series converges.

b. We can see that

$$\rho = \lim_{n\to\infty} \frac{(n+1)^{n+1}/(n+1)!}{n^n/n!} = \lim_{n\to\infty} \frac{(n+1)^{n+1}}{(n+1)!} \cdot \frac{n!}{n^n} = \lim_{n\to\infty} \left(\frac{n+1}{n}\right)^n = \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = e.$$

Since $\rho > 1$, the series diverges.

c. Since

$$\left| \frac{(-1)^{n+1}((n+1)!)^2/(2(n+1))!}{(-1)^n (n!)^2/(2n)!} \right| = \frac{(n+1)!(n+1)!}{(2n+2)!} \cdot \frac{(2n)!}{n!n!} = \frac{(n+1)(n+1)}{(2n+2)(2n+1)}$$

we see that

$$\rho = \lim_{n\to\infty} \frac{(n+1)(n+1)}{(2n+2)(2n+1)} = \frac{1}{4}.$$

Since $\rho < 1$, the series converges.

Exercise 5.6.1

Use the ratio test to determine whether the series $\sum_{n=1}^{\infty} \frac{n^3}{3^n}$ converges or diverges.

Hint

Evaluate $\lim_{n\to\infty} \dfrac{(n+1)^3}{3^{n+1}} \cdot \dfrac{3^n}{n^3}$.

Answer

The series converges.

Root Test

The approach of the **root test** is similar to that of the ratio test. Consider a series $\sum_{n=1}^{\infty} a_n$ such that $\lim_{n\to\infty} \sqrt[n]{|a_n|} = \rho$ for some real number ρ. Then for N sufficiently large, $|a_N| \approx \rho^N$. Therefore, we can approximate $\sum_{n=N}^{\infty} |a_n|$ by writing

$$|a_N| + |a_{N+1}| + |a_{N+2}| + \cdots \approx \rho^N + \rho^{N+1} + \rho^{N+2} + \cdots.$$

The expression on the right-hand side is a geometric series. As in the ratio test, the series $\sum_{n=1}^{\infty} a_n$ converges absolutely if $0 \le \rho < 1$ and the series diverges if $\rho \ge 1$. If $\rho = 1$, the test does not provide any information. For example, for any p-series, $\sum_{n=1}^{\infty} \dfrac{1}{n^p}$, we see that

$$\rho = \lim_{n\to\infty} \sqrt[n]{\left|\frac{1}{n^p}\right|} = \lim_{n\to\infty} \frac{1}{n^{p/n}}$$

.

To evaluate this limit, we use the natural logarithm function. Doing so, we see that

$$\ln \rho = \ln\left(\lim_{n\to\infty} \frac{1}{n^{p/n}}\right) = \lim_{n\to\infty} \ln\left(\frac{1}{n}\right)^{p/n} = \lim_{n\to\infty} \frac{p}{n} \cdot \ln\left(\frac{1}{n}\right) = \lim_{n\to\infty} \frac{p\ln(1/n)}{n}.$$

Using L'Hôpital's rule, it follows that $\ln \rho = 0$, and therefore $\rho = 1$ for all p. However, we know that the p-series only converges if $p > 1$ and diverges if $p < 1$.

Root Test

Consider the series $\sum_{n=1}^{\infty} a_n$. Let

$$\rho = \lim_{n\to\infty} \sqrt[n]{|a_n|}$$

.

i. If $0 \le \rho < 1$, then $\sum_{n=1}^{\infty} a_n$ converges absolutely.

ii. If $\rho > 1$ or $\rho = \infty$, then $\sum_{n=1}^{\infty} a_n$ diverges.

iii. If $\rho = 1$, the test does not provide any information.

The root test is useful for series whose terms involve exponentials. In particular, for a series whose terms a_n satisfy $|a_n| = (b_n)^n$, then $\sqrt[n]{|a_n|} = b_n$ and we need only evaluate $\lim_{n\to\infty} b_n$.

Example 5.6.2: Using the Root Test

For each of the following series, use the root test to determine whether the series converges or diverges.

a. $\displaystyle\sum_{n=1}^{\infty}\frac{(n^2+3n)^n}{(4n^2+5)^n}$

b. $\displaystyle\sum_{n=1}^{\infty}\frac{n^n}{(\ln(n))^n}$

Solution

a. To apply the root test, we compute

$$\rho=\lim_{n\to\infty}\sqrt[n]{\frac{(n^2+3n)^n}{(4n^2+5)^n}}=\lim_{n\to\infty}\frac{n^2+3n}{4n^2+5}=\frac{1}{4}.$$

Since $\rho<1$, the series converges absolutely.

b. We have

$$\rho=\lim_{n\to\infty}\sqrt[n]{\frac{n^n}{(\ln(n))^n}}=\lim_{n\to\infty}\frac{n}{\ln n}=\infty \quad \text{by L'Hôpital's rule.}$$

Since $\rho=\infty$, the series diverges.

Exercise 5.6.2

Use the root test to determine whether the series $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^n}$ converges or diverges.

Hint

Evaluate $\displaystyle\lim_{n\to\infty}\sqrt[n]{\frac{1}{n^n}}$.

Answer

The series converges.

Choosing a Convergence Test

At this point, we have a long list of convergence tests. However, not all tests can be used for all series. When given a series, we must determine which test is the best to use. Here is a strategy for finding the best test to apply.

Problem-Solving Strategy: Choosing a Convergence Test for a Series

Consider a series $\displaystyle\sum_{n=1}^{\infty}a_n$. In the steps below, we outline a strategy for determining whether the series converges.

1. Is $\displaystyle\sum_{n=1}^{\infty}a_n$ a familiar series? For example, is it the harmonic series (which diverges) or the alternating harmonic series (which converges)? Is it a **p−series** or geometric series? If so, check the power p or the ratio r to determine if the series converges.
2. Is it an alternating series? Are we interested in absolute convergence or just convergence? If we are just interested in whether the series converges, apply the alternating series test. If we are interested in absolute convergence, proceed to step 3, considering the series of absolute values $\displaystyle\sum_{n=1}^{\infty}|a_n|$.
3. Is the series similar to a **p−series** or geometric series? If so, try the comparison test or limit comparison test.

4. Do the terms in the series contain a factorial or power? If the terms are powers such that $a_n = (b_n)^n$, try the root test first. Otherwise, try the ratio test first.
5. Use the divergence test. If this test does not provide any information, try the integral test.

Example 5.6.3: Using Convergence Tests

For each of the following series, determine which convergence test is the best to use and explain why. Then determine if the series converges or diverges. If the series is an alternating series, determine whether it converges absolutely, converges conditionally, or diverges.

a. $\displaystyle\sum_{n=1}^{\infty} \frac{n^2+2n}{n^3+3n^2+1}$

b. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}(3n+1)}{n!}$

c. $\displaystyle\sum_{n=1}^{\infty} \frac{e^n}{n^3}$

d. $\displaystyle\sum_{n=1}^{\infty} \frac{3^n}{(n+1)^n}$

Solution

a. Step 1. The series is not a **p–series** or geometric series.

Step 2. The series is not alternating.

Step 3. For large values of n, we approximate the series by the expression

$$\frac{n^2+2n}{n^3+3n^2+1} \approx \frac{n^2}{n^3} = \frac{1}{n}.$$

Therefore, it seems reasonable to apply the comparison test or limit comparison test using the series $\displaystyle\sum_{n=1}^{\infty} 1/n$.

Using the limit comparison test, we see that

$$\lim_{n\to\infty} \frac{(n^2+2n)/(n^3+3n^2+1)}{1/n} = \lim_{n\to\infty} \frac{n^3+2n^2}{n^3+3n^2+1} = 1.$$

Since the series $\displaystyle\sum_{n=1}^{\infty} 1/n$

diverges, this series diverges as well.

b. Step 1.The series is not a familiar series.

Step 2. The series is alternating. Since we are interested in absolute convergence, consider the series

$$\sum_{n=1}^{\infty} \frac{3n}{(n+1)!}.$$

Step 3. The series is not similar to a p-series or geometric series.

Step 4. Since each term contains a factorial, apply the ratio test. We see that

$$\lim_{n\to\infty} \frac{(3(n+1))/(n+1)!}{(3n+1)/n!} = \lim_{n\to\infty} \frac{3n+3}{(n+1)!} \cdot \frac{n!}{3n+1} = \lim_{n\to\infty} \frac{3n+3}{(n+1)(3n+1)} = 0.$$

Therefore, this series converges, and we conclude that the original series converges absolutely, and thus converges.

c. Step 1. The series is not a familiar series.

Step 2. It is not an alternating series.

Step 3. There is no obvious series with which to compare this series.

Step 4. There is no factorial. There is a power, but it is not an ideal situation for the root test.

Step 5. To apply the divergence test, we calculate that

$$\lim_{n\to\infty}\frac{e^n}{n^3}=\infty.$$

Therefore, by the divergence test, the series diverges.

d. Step 1. This series is not a familiar series.

Step 2. It is not an alternating series.

Step 3. There is no obvious series with which to compare this series.

Step 4. Since each term is a power of n,we can apply the root test. Since

$$\lim_{n\to\infty}\sqrt[n]{\left(\frac{3}{n+1}\right)^n}=\lim_{n\to\infty}\frac{3}{n+1}=0,$$

by the root test, we conclude that the series converges.

? Exercise 5.6.3

For the series $\sum_{n=1}^{\infty}\frac{2^n}{3^n+n}$, determine which convergence test is the best to use and explain why.

Hint

The series is similar to the geometric series $\sum_{n=1}^{\infty}\left(\frac{2}{3}\right)^n$.

Answer

The comparison test because $\frac{2^n}{3^n+n}<\frac{2^n}{3^n}$ for all positive integers n. The limit comparison test could also be used.

In Table, we summarize the convergence tests and when each can be applied. Note that while the comparison test, limit comparison test, and integral test require the series $\sum_{n=1}^{\infty}a_n$ to have nonnegative terms, if $\sum_{n=1}^{\infty}a_n$ has negative terms, these tests can be applied to $\sum_{n=1}^{\infty}|a_n|$ to test for absolute convergence.

Summary of Convergence Tests

Series or Test	Conclusions	Comments
Divergence Test For any series $\sum_{n=1}^{\infty}a_n$, evaluate $\lim_{n\to\infty}a_n$.	If $\lim_{n\to\infty}a_n=0$, the test is inconclusive.	This test cannot prove convergence of a series.
	If $\lim_{n\to\infty}a_n\neq 0$, the series diverges.	
Geometric Series $\sum_{n=1}^{\infty}ar^{n-1}$	If $\lvert r\rvert<1$, the series converges to $a/(1-r)$.	Any geometric series can be reindexed to be written in the form $a+ar+ar^2+\cdots$, where a is the initial term and r is the ratio.
	If $\lvert r\rvert\geq 1$, the series diverges.	
p-Series $\sum_{n=1}^{\infty}\frac{1}{n^p}$	If $p>1$, the series converges.	For $p=1$, we have the harmonic series $\sum_{n=1}^{\infty}1/n$.
	If $p\leq 1$, the series diverges.	

Series or Test	Conclusions	Comments
Comparison Test For $\sum_{n=1}^{\infty} a_n$ with nonnegative terms, compare with a known series $\sum_{n=1}^{\infty} b_n$.	If $a_n \le b_n$ for all $n \ge N$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.	Typically used for a series similar to a geometric or p-series. It can sometimes be difficult to find an appropriate series.
	If $a_n \ge b_n$ for all $n \ge N$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.	
Limit Comparison Test For $\sum_{n=1}^{\infty} a_n$ with positive terms, compare with a series $\sum_{n=1}^{\infty} b_n$ by evaluating $L = \lim_{n\to\infty} \frac{a_n}{b_n}$.	If L is a real number and $L \ne 0$, then $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge or both diverge.	Typically used for a series similar to a geometric or p-series. Often easier to apply than the comparison test.
	If $L = 0$ and $\sum_{n=1}^{\infty} b_n$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.	
	If $L = \infty$ and $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.	
Integral Test If there exists a positive, continuous, decreasing function f such that $a_n = f(n)$ for all $n \ge N$, evaluate $\int_N^{\infty} f(x)dx$.	$\int_N^{\infty} f(x)dx$ and $\sum_{n=1}^{\infty} a_n$ both converge or both diverge.	Limited to those series for which the corresponding function can be easily integrated.
Alternating Series $\sum_{n=1}^{\infty}(-1)^{n+1}b_n$ or $\sum_{n=1}^{\infty}(-1)^{n}b_n$	If $b_{n+1} \le b_n$ for all $n \ge 1$ and $b_n \to 0$, then the series converges.	Only applies to alternating series.
Ratio Test For any series $\sum_{n=1}^{\infty} a_n$ with nonzero terms, let $\rho = \lim_{n\to\infty} \left\lvert \frac{a_{n+1}}{a_n} \right\rvert$	If $0 \le \rho < 1$, the series converges absolutely.	Often used for series involving factorials or exponentials.
	If $\rho > 1$ or $\rho = \infty$, the series diverges.	
	If $\rho = 1$, the test is inconclusive.	
Root Test For any series $\sum_{n=1}^{\infty} a_n$, let $\rho = \lim_{n\to\infty} \sqrt[n]{\lvert a_n \rvert}$.	If $0 \le \rho < 1$, the series converges absolutely.	Often used for series where $\lvert a_n \rvert = (b_n)^n$.
	If $\rho > 1$ or $\rho = \infty$, the series diverges.	
	If $\rho = 1$, the test is inconclusive.	

Series Converging to π and $1/\pi$

Dozens of series exist that converge to π or an algebraic expression containing π. Here we look at several examples and compare their rates of convergence. By rate of convergence, we mean the number of terms necessary for a partial sum to be within a certain amount of the actual value. The series representations of π in the first two examples can be explained using Maclaurin series, which are discussed in the next chapter. The third example relies on material beyond the scope of this text.

1. The series

$$\pi = 4\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n-1} = 4 - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \cdots$$

was discovered by Gregory and Leibniz in the late 1600s. This result follows from the Maclaurin series for $f(x) = \tan^{-1} x$. We will discuss this series in the next chapter.

a. Prove that this series converges.

b. Evaluate the partial sums S_n for $n = 10, 20, 50, 100$.

c. Use the remainder estimate for alternating series to get a bound on the error R_n.

d. What is the smallest value of N that guarantees $|R_N| < 0.01$? Evaluate S_N.

2. The series

$$\pi = 6\sum_{n=0}^{\infty}\frac{(2n)!}{2^{4n+1}(n!)^2(2n+1)} = 6\left(\frac{1}{2} + \frac{1}{2\cdot 3}\left(\frac{1}{2}\right)^3 + \frac{1\cdot 3}{2\cdot 4\cdot 5}\cdot\left(\frac{1}{2}\right)^5 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 7}\left(\frac{1}{2}\right)^7 + \cdots\right)$$

has been attributed to Newton in the late 1600s. The proof of this result uses the Maclaurin series for $f(x) = \sin^{-1} x$.

a. Prove that the series converges.

b. Evaluate the partial sums S_n for $n = 5, 10, 20$.

c. Compare S_n to π for $n = 5, 10, 20$ and discuss the number of correct decimal places.

3. The series

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801}\sum_{n=0}^{\infty}\frac{(4n)!(1103 + 26390n)}{(n!)^4 396^{4n}}$$

was discovered by **Ramanujan** in the early 1900s. William Gosper, Jr., used this series to calculate π to an accuracy of more than 17 million digits in the mid-1980s. At the time, that was a world record. Since that time, this series and others by Ramanujan have led mathematicians to find many other series representations for π and $1/\pi$.

a. Prove that this series converges.

b. Evaluate the first term in this series. Compare this number with the value of π from a calculating utility. To how many decimal places do these two numbers agree? What if we add the first two terms in the series?

c. Investigate the life of Srinivasa Ramanujan (1887–1920) and write a brief summary. Ramanujan is one of the most fascinating stories in the history of mathematics. He was basically self-taught, with no formal training in mathematics, yet he contributed in highly original ways to many advanced areas of mathematics.

Key Concepts

- For the ratio test, we consider

$$\rho = \lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right|.$$

If $\rho < 1$, the series $\sum_{n=1}^{\infty} a_n$ converges absolutely. If $\rho > 1$, the series diverges. If $\rho = 1$, the test does not provide any information. This test is useful for series whose terms involve factorials.

- For the root test, we consider

$$\rho = \lim_{n\to\infty}\sqrt[n]{|a_n|}.$$

If $\rho < 1$, the series $\sum_{n=1}^{\infty} a_n$ converges absolutely. If $\rho > 1$, the series diverges. If $\rho = 1$, the test does not provide any information. The root test is useful for series whose terms involve powers.

- For a series that is similar to a geometric series or **p−series,** consider one of the comparison tests.

Glossary

ratio test

for a series $\sum_{n=1}^{\infty} a_n$ with nonzero terms, let $\rho = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$; if $0 \le \rho < 1$, the series converges absolutely; if $\rho > 1$, the series diverges; if $\rho = 1$, the test is inconclusive

root test

for a series $\sum_{n=1}^{\infty} a_n$, let $\rho = \lim_{n\to\infty} \sqrt[n]{|a_n|}$; if $0 \le \rho < 1$, the series converges absolutely; if $\rho > 1$, the series diverges; if $\rho = 1$, the test is inconclusive

---

This page titled 5.6: Ratio and Root Tests is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **9.6: Ratio and Root Tests** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

CHAPTER OVERVIEW

6: Power Series

A power series (in one variable) is an infinite series. Any polynomial can be easily expressed as a power series around any center c, although most of the coefficients will be zero since a power series has infinitely many terms by definition. One can view power series as being like "polynomials of infinite degree," although power series are not polynomials. The content in this Textmap's chapter is complemented by Guichard's Calculus Textmap.

6.E: Power Series (Exercises)
6.0: Prelude to Power Series
6.1: Power Series and Functions
6.2: Properties of Power Series
6.3: Taylor and Maclaurin Series
6.4: Working with Taylor Series

Thumbnail: The graph shows the function $y = sinx$ and the Maclaurin polynomials p_1, p_3 and p_5. Image used with permission (CC BY-SA 3.0; OpenStax).

Contributors

Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 6: Power Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

6.E: Power Series (Exercises)

10.E: Power Series (Exercises)

This page titled 6.E: Power Series (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

6.0: Prelude to Power Series

When winning a lottery, sometimes an individual has an option of receiving winnings in one lump-sum payment or receiving smaller payments over fixed time intervals. For example, you might have the option of receiving 20 million dollars today or receiving 1.5 million dollars each year for the next 20 years. Which is the better deal? Certainly 1.5 million dollars over 20 years is equivalent to 30 million dollars. However, receiving the 20 million dollars today would allow you to invest the money.



Figure 6.0.1: If you win a lottery, do you get more money by taking a lump-sum payment or by accepting fixed payments over time? (credit: modification of work by Robert Huffstutter, Flickr)

Alternatively, what if you were guaranteed to receive 1 million dollars every year indefinitely (extending to your heirs) or receive 20 million dollars today. Which would be the better deal? To answer these questions, you need to know how to use infinite series to calculate the value of periodic payments over time in terms of today's dollars.

An infinite series of the form

$$\sum_{n=0}^{\infty} c_n x^n$$

is known as a **power series**. Since the terms contain the variable x, power series can be used to define functions. They can be used to represent given functions, but they are also important because they allow us to write functions that cannot be expressed any other way than as "infinite polynomials." In addition, power series can be easily differentiated and integrated, thus being useful in solving differential equations and integrating complicated functions. An infinite series can also be truncated, resulting in a finite polynomial that we can use to approximate functional values. Power series have applications in a variety of fields, including physics, chemistry, biology, and economics. As we will see in this chapter, representing functions using power series allows us to solve mathematical problems that cannot be solved with other techniques.

Gilbert Strang (MIT) and Edwin "Jed" Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 6.0: Prelude to Power Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.0: Prelude to Power Series** by Edwin "Jed" Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

6.1: Power Series and Functions

Learning Objectives

- Identify a power series and provide examples of them.
- Determine the radius of convergence and interval of convergence of a power series.
- Use a power series to represent a function.

A power series is a type of series with terms involving a variable. More specifically, if the variable is *[Math Processing Error]*, then all the terms of the series involve powers of *[Math Processing Error]*. As a result, a power series can be thought of as an infinite polynomial. Power series are used to represent common functions and also to define new functions. In this section we define power series and show how to determine when a power series converges and when it diverges. We also show how to represent certain functions using power series.

Form of a Power Series

A series of the form

[Math Processing Error]

where *[Math Processing Error]* is a variable and the coefficients *[Math Processing Error]* are constants, is known as a **power series.** The series

[Math Processing Error]

is an example of a power series. Since this series is a geometric series with ratio *[Math Processing Error]*, we know that it converges if *[Math Processing Error]* and diverges if *[Math Processing Error]*

Definition *[Math Processing Error]*: Power Series

A series of the form

[Math Processing Error]

is a power series centered at *[Math Processing Error]* A series of the form

[Math Processing Error]

is a *power series* centered at *[Math Processing Error]*.

To make this definition precise, we stipulate that *[Math Processing Error]* and *[Math Processing Error]* even when *[Math Processing Error]* and *[Math Processing Error]*, respectively.

The series

[Math Processing Error]

and

[Math Processing Error]

are both power series centered at *[Math Processing Error]* The series

[Math Processing Error]

is a power series centered at *[Math Processing Error]*.

Convergence of a Power Series

Since the terms in a power series involve a variable *[Math Processing Error]*, the series may converge for certain values of *[Math Processing Error]* and diverge for other values of *[Math Processing Error]*. For a power series centered at *[Math Processing Error]*, the value of the series at *[Math Processing Error]* is given by *[Math Processing Error]*. Therefore, a power series always converges at its center. Some power series converge only at that value of *[Math Processing Error]*. Most power series, however, converge for more than one value of *[Math Processing Error]*. In that case, the power series either converges for all real numbers

[Math Processing Error] or converges for all [Math Processing Error] in a finite interval. For example, the geometric series [Math Processing Error] converges for all [Math Processing Error] in the interval [Math Processing Error], but diverges for all [Math Processing Error] outside that interval. We now summarize these three possibilities for a general power series.

Note [Math Processing Error]: Convergence of a Power Series

Consider the power series [Math Processing Error] The series satisfies exactly one of the following properties:

i. The series converges at [Math Processing Error] and diverges for all [Math Processing Error]
ii. The series converges for all real numbers [Math Processing Error].
iii. There exists a real number [Math Processing Error] such that the series converges if [Math Processing Error] and diverges if [Math Processing Error]. At the values [Math Processing Error] where [Math Processing Error], the series may converge or diverge.

Proof

Suppose that the power series is centered at [Math Processing Error]. (For a series centered at a value of [Math Processing Error] other than zero, the result follows by letting [Math Processing Error] and considering the series

[Math Processing Error]

We must first prove the following fact:

If there exists a real number [Math Processing Error] such that [Math Processing Error] converges, then the series [Math Processing Error] converges absolutely for all [Math Processing Error] such that [Math Processing Error]

Since [Math Processing Error] converges, the nth term [Math Processing Error] as [Math Processing Error]. Therefore, there exists an integer [Math Processing Error] such that [Math Processing Error] for all [Math Processing Error] Writing

[Math Processing Error]

we conclude that, for all n≥N,

[Math Processing Error]

The series

[Math Processing Error]

is a geometric series that converges if [Math Processing Error] Therefore, by the comparison test, we conclude that [Math Processing Error] also converges for [Math Processing Error]. Since we can add a finite number of terms to a convergent series, we conclude that [Math Processing Error] converges for [Math Processing Error]

With this result, we can now prove the theorem. Consider the series

[Math Processing Error]

and let [Math Processing Error] be the set of real numbers for which the series converges. Suppose that the set [Math Processing Error] Then the series falls under case i.

Suppose that the set [Math Processing Error] is the set of all real numbers. Then the series falls under case ii. Suppose that [Math Processing Error] and [Math Processing Error] is not the set of real numbers. Then there exists a real number [Math Processing Error] such that the series does not converge. Thus, the series cannot converge for any [Math Processing Error] such that [Math Processing Error]. Therefore, the set [Math Processing Error] must be a bounded set, which means that it must have a smallest upper bound. (This fact follows from the **Least Upper Bound Property** for the real numbers, which is beyond the scope of this text and is covered in real analysis courses.) Call that smallest upper bound [Math Processing Error]. Since [Math Processing Error], the number [Math Processing Error]. Therefore, the series converges for all [Math Processing Error] such that [Math Processing Error] and the series falls into case iii.

□

If a series [Math Processing Error] falls into case iii. of Note [Math Processing Error], then the series converges for all [Math Processing Error] such that [Math Processing Error] for some [Math Processing Error], and diverges for all [Math Processing

Error] such that *[Math Processing Error]*. The series may converge or diverge at the values *[Math Processing Error]* where *[Math Processing Error]*. The set of values *[Math Processing Error]* for which the series *[Math Processing Error]* converges is known as the interval of convergence. Since the series diverges for all values *[Math Processing Error]* where *[Math Processing Error]*, the length of the interval is *[Math Processing Error]*, and therefore, the radius of the interval is *[Math Processing Error]*. The value *[Math Processing Error]* is called the radius of convergence. For example, since the series *[Math Processing Error]* converges for all values *[Math Processing Error]* in the interval *[Math Processing Error]* and diverges for all values *[Math Processing Error]* such that *[Math Processing Error]*, the interval of convergence of this series is *[Math Processing Error]*. Since the length of the interval is *[Math Processing Error]*, the radius of convergence is *[Math Processing Error]*.

Definition: Radius of Convergence

Consider the power series *[Math Processing Error]*. The set of real numbers *[Math Processing Error]* where the series converges is the interval of convergence. If there exists a real number *[Math Processing Error]* such that the series converges for *[Math Processing Error]* and diverges for *[Math Processing Error]* then *[Math Processing Error]* is the radius of convergence. If the series converges only at *[Math Processing Error]*, we say the radius of convergence is *[Math Processing Error]*. If the series converges for all real numbers *[Math Processing Error]*, we say the radius of convergence is *[Math Processing Error]* (Figure *[Math Processing Error]*).

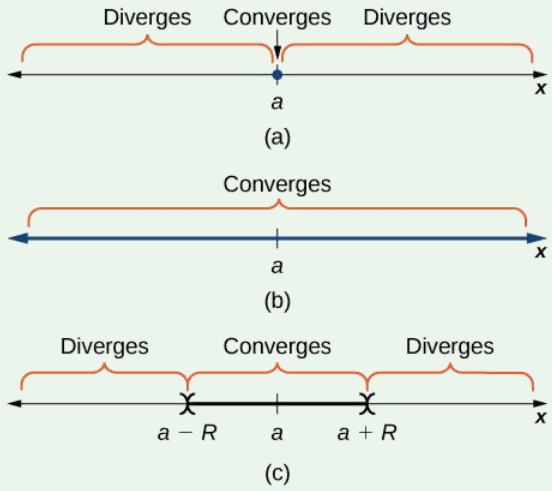


Figure *[Math Processing Error]*: For a series *[Math Processing Error]* graph (a) shows a radius of convergence at *[Math Processing Error]*, graph (b) shows a radius of convergence at *[Math Processing Error]*, and graph (c) shows a radius of convergence at *[Math Processing Error]*. For graph (c) we note that the series may or may not converge at the endpoints *[Math Processing Error]* and *[Math Processing Error]*.

To determine the interval of convergence for a power series, we typically apply the ratio test. In Example *[Math Processing Error]*, we show the three different possibilities illustrated in Figure *[Math Processing Error]*.

Example *[Math Processing Error]*: Finding the Interval and Radius of Convergence

For each of the following series, find the interval and radius of convergence.

a. *[Math Processing Error]*
b. *[Math Processing Error]*
c. *[Math Processing Error]*

Solution

a. To check for convergence, apply the ratio test. We have

[Math Processing Error]

for all values of *[Math Processing Error]*. Therefore, the series converges for all real numbers *[Math Processing Error]*. The interval of convergence is *[Math Processing Error]* and the radius of convergence is *[Math Processing Error]*

b. Apply the ratio test. For *[Math Processing Error]*, we see that

[Math Processing Error]

Therefore, the series diverges for all [Math Processing Error]. Since the series is centered at [Math Processing Error], it must converge there, so the series converges only for [Math Processing Error]. The interval of convergence is the single value [Math Processing Error] and the radius of convergence is [Math Processing Error].

c. In order to apply the ratio test, consider

[Math Processing Error]

The ratio [Math Processing Error] if [Math Processing Error]. Since [Math Processing Error] implies that [Math Processing Error] the series converges absolutely if [Math Processing Error]. The ratio [Math Processing Error] if [Math Processing Error]. Therefore, the series diverges if [Math Processing Error] or [Math Processing Error]. The ratio test is inconclusive if [Math Processing Error]. The ratio [Math Processing Error] if and only if [Math Processing Error] or [Math Processing Error]. We need to test these values of [Math Processing Error] separately. For [Math Processing Error], the series is given by

[Math Processing Error]

Since this is the alternating harmonic series, it converges. Thus, the series converges at [Math Processing Error]. For [Math Processing Error], the series is given by

[Math Processing Error]

This is the harmonic series, which is divergent. Therefore, the power series diverges at [Math Processing Error]. We conclude that the interval of convergence is [Math Processing Error] and the radius of convergence is [Math Processing Error].

? Exercise [Math Processing Error]

Find the interval and radius of convergence for the series

[Math Processing Error]

Hint

Apply the ratio test to check for absolute convergence.

Answer

The interval of convergence is [Math Processing Error] The radius of convergence is [Math Processing Error]

Representing Functions as Power Series

Being able to represent a function by an “infinite polynomial” is a powerful tool. Polynomial functions are the easiest functions to analyze, since they only involve the basic arithmetic operations of addition, subtraction, multiplication, and division. If we can represent a complicated function by an infinite polynomial, we can use the polynomial representation to differentiate or integrate it. In addition, we can use a truncated version of the polynomial expression to approximate values of the function. So, the question is, when can we represent a function by a power series?

Consider again the geometric series

[Math Processing Error]

Recall that the geometric series

[Math Processing Error]

converges if and only if [Math Processing Error] In that case, it converges to [Math Processing Error]. Therefore, if [Math Processing Error], the series in Example [Math Processing Error] converges to [Math Processing Error] and we write

[Math Processing Error]

As a result, we are able to represent the function [Math Processing Error] by the power series

[Math Processing Error]

We now show graphically how this series provides a representation for the function *[Math Processing Error]* by comparing the graph of *[Math Processing Error]* with the graphs of several of the partial sums of this infinite series.

Example *[Math Processing Error]*: Graphing a Function and Partial Sums of its Power Series

Sketch a graph of *[Math Processing Error]* and the graphs of the corresponding partial sums *[Math Processing Error]* for *[Math Processing Error]* on the interval *[Math Processing Error]*. Comment on the approximation *[Math Processing Error]* as *[Math Processing Error]* increases.

Solution

From the graph in Figure *[Math Processing Error]*, you see that as *[Math Processing Error]* increases, *[Math Processing Error]* becomes a better approximation for *[Math Processing Error]* for *[Math Processing Error]* in the interval *[Math Processing Error]*.

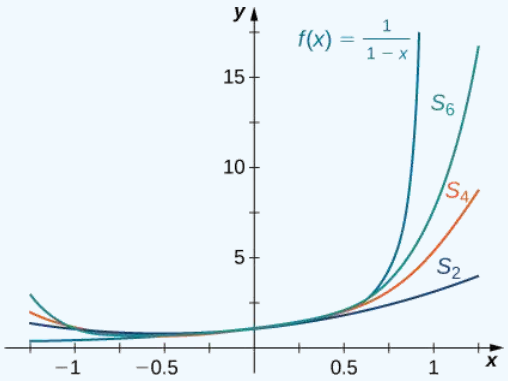


Figure *[Math Processing Error]*: The graph shows a function and three approximations of it by partial sums of a power series.

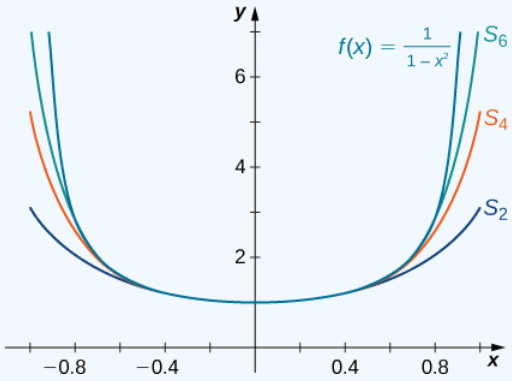
Exercise *[Math Processing Error]*

Sketch a graph of *[Math Processing Error]* and the corresponding partial sums *[Math Processing Error]* for *[Math Processing Error]* on the interval *[Math Processing Error]*

Hint

[Math Processing Error]

Answer



Next we consider functions involving an expression similar to the sum of a geometric series and show how to represent these functions using power series.

Example *[Math Processing Error]*: Representing a Function with a Power Series

Use a power series to represent each of the following functions *[Math Processing Error]*. Find the interval of convergence.

a. *[Math Processing Error]*
b. *[Math Processing Error]*

Solution

a. You should recognize this function *[Math Processing Error]* as the sum of a geometric series, because

[Math Processing Error]

Using the fact that, for *[Math Processing Error]* is the sum of the geometric series

[Math Processing Error]

we see that, for *[Math Processing Error]*

[Math Processing Error]

Since this series converges if and only if *[Math Processing Error]*, the interval of convergence is *[Math Processing Error]*, and we have

[Math Processing Error]

b. This function is not in the exact form of a sum of a geometric series. However, with a little algebraic manipulation, we can relate f to a geometric series. By factoring 4 out of the two terms in the denominator, we obtain

[Math Processing Error]

Therefore, we have

[Math Processing Error]

The series converges as long as *[Math Processing Error]* (note that when *[Math Processing Error]* the series does not converge). Solving this inequality, we conclude that the interval of convergence is *[Math Processing Error]* and

[Math Processing Error]

for *[Math Processing Error]*

Exercise *[Math Processing Error]*

Represent the function *[Math Processing Error]* using a power series and find the interval of convergence.

Hint

Rewrite f in the form *[Math Processing Error]* for some functions *[Math Processing Error]* and *[Math Processing Error]*.

Answer

[Math Processing Error] with interval of convergence *[Math Processing Error]*

In the remaining sections of this chapter, we will show ways of deriving power series representations for many other functions, and how we can make use of these representations to evaluate, differentiate, and integrate various functions.

Key Concepts

- For a power series centered at *[Math Processing Error]*, one of the following three properties hold:
  - i. The power series converges only at *[Math Processing Error]*. In this case, we say that the radius of convergence is *[Math Processing Error]*.
  - ii. The power series converges for all real numbers *[Math Processing Error]*. In this case, we say that the radius of convergence is *[Math Processing Error]*.
  - iii. There is a real number R such that the series converges for *[Math Processing Error]* and diverges for *[Math Processing Error]*. In this case, the radius of convergence is *[Math Processing Error]*

- If a power series converges on a finite interval, the series may or may not converge at the endpoints.
- The ratio test may often be used to determine the radius of convergence.
- The geometric series *[Math Processing Error]* for *[Math Processing Error]* allows us to represent certain functions using geometric series.

Key Equations

- **Power series centered at *[Math Processing Error]***

  [Math Processing Error]

- **Power series centered at *[Math Processing Error]***

  [Math Processing Error]

Glossary

interval of convergence

the set of real numbers *[Math Processing Error]* for which a power series converges

power series

a series of the form *[Math Processing Error]* is a power series centered at *[Math Processing Error]*; a series of the form *[Math Processing Error]* is a power series centered at *[Math Processing Error]*

radius of convergence

if there exists a real number *[Math Processing Error]* such that a power series centered at *[Math Processing Error]* converges for *[Math Processing Error]* and diverges for *[Math Processing Error]*, then *[Math Processing Error]* is the radius of convergence; if the power series only converges at *[Math Processing Error]*, the radius of convergence is *[Math Processing Error]*; if the power series converges for all real numbers *[Math Processing Error]*, the radius of convergence is *[Math Processing Error]*

This page titled 6.1: Power Series and Functions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Properties of Power Series

Learning Objectives

- Combine power series by addition or subtraction.
- Create a new power series by multiplication by a power of the variable or a constant, or by substitution.
- Multiply two power series together.
- Differentiate and integrate power series term-by-term.

In the preceding section on power series and functions we showed how to represent certain functions using power series. In this section we discuss how power series can be combined, differentiated, or integrated to create new power series. This capability is particularly useful for a couple of reasons. First, it allows us to find power series representations for certain elementary functions, by writing those functions in terms of functions with known power series. For example, given the power series representation for *[Math Processing Error]*, we can find a power series representation for *[Math Processing Error]*. Second, being able to create power series allows us to define new functions that cannot be written in terms of elementary functions. This capability is particularly useful for solving differential equations for which there is no solution in terms of elementary functions.

Combining Power Series

If we have two power series with the same interval of convergence, we can add or subtract the two series to create a new power series, also with the same interval of convergence. Similarly, we can multiply a power series by a power of *[Math Processing Error]* or evaluate a power series at *[Math Processing Error]* for a positive integer *[Math Processing Error]* to create a new power series. Being able to do this allows us to find power series representations for certain functions by using power series representations of other functions. For example, since we know the power series representation for *[Math Processing Error]*, we can find power series representations for related functions, such as

[Math Processing Error]

and

[Math Processing Error]

In Theorem *[Math Processing Error]*, we state results regarding addition or subtraction of power series, composition of a power series, and multiplication of a power series by a power of the variable. For simplicity, we state the theorem for power series centered at *[Math Processing Error]*. Similar results hold for power series centered at *[Math Processing Error]*.

Theorem *[Math Processing Error]*: Combining Power Series

Suppose that the two power series *[Math Processing Error]* and *[Math Processing Error]* converge to the functions *[Math Processing Error]* and *[Math Processing Error]*, respectively, on a common interval *[Math Processing Error]*.

i. The power series *[Math Processing Error]* converges to *[Math Processing Error]* on *[Math Processing Error]*.
ii. For any integer *[Math Processing Error]* and any real number *[Math Processing Error]*, the power series *[Math Processing Error]* converges to *[Math Processing Error]* on *[Math Processing Error]*.
iii. For any integer *[Math Processing Error]* and any real number *[Math Processing Error]*, the series *[Math Processing Error]* converges to *[Math Processing Error]* for all *[Math Processing Error]* such that *[Math Processing Error]* is in *[Math Processing Error]*.

Proof

We prove *[Math Processing Error]*. in the case of the series *[Math Processing Error]*. Suppose that *[Math Processing Error]* and *[Math Processing Error]* converge to the functions *[Math Processing Error]* and *[Math Processing Error]*, respectively, on the interval *[Math Processing Error]*. Let *[Math Processing Error]* be a point in *[Math Processing Error]* and let *[Math Processing Error]* and *[Math Processing Error]* denote the *[Math Processing Error]* partial sums of the series *[Math Processing Error]* and *[Math Processing Error]*, respectively. Then the sequence *[Math Processing Error]* converges to *[Math Processing Error]* and the sequence *[Math Processing Error]* converges to *[Math Processing Error]*. Furthermore, the *[Math Processing Error]* partial sum of *[Math Processing Error]* is

[Math Processing Error]

Because

[Math Processing Error]

we conclude that the series *[Math Processing Error]* converges to *[Math Processing Error]*

□

We examine products of power series in a later theorem. First, we show several applications of Theorem *[Math Processing Error]* and how to find the interval of convergence of a power series given the interval of convergence of a related power series.

✔ Example *[Math Processing Error]*: Combining Power Series

Suppose that *[Math Processing Error]* is a power series whose interval of convergence is *[Math Processing Error]*, and suppose that *[Math Processing Error]* is a power series whose interval of convergence is *[Math Processing Error]*

a. Find the interval of convergence of the series *[Math Processing Error]*
b. Find the interval of convergence of the series *[Math Processing Error]*

Solution

a. Since the interval *[Math Processing Error]* is a common interval of convergence of the series *[Math Processing Error]* and *[Math Processing Error]*, the interval of convergence of the series *[Math Processing Error]* is *[Math Processing Error]*.
b. Since *[Math Processing Error]* is a power series centered at zero with radius of convergence *[Math Processing Error]* it converges for all *[Math Processing Error]* in the interval *[Math Processing Error]* By Theorem *[Math Processing Error]*, the series *[Math Processing Error]* converges if *[Math Processing Error]* is in the interval *[Math Processing Error]*. Therefore, the series converges for all *[Math Processing Error]* in the interval *[Math Processing Error]*

? Exercise *[Math Processing Error]*

Suppose that *[Math Processing Error]* has an interval of convergence of *[Math Processing Error]*. Find the interval of convergence of *[Math Processing Error]*.

Hint

Find the values of *[Math Processing Error]* such that *[Math Processing Error]* is in the interval *[Math Processing Error]*

Answer

Interval of convergence is *[Math Processing Error]*

In the next example, we show how to use Theorem *[Math Processing Error]* and the power series for a function *[Math Processing Error]* to construct power series for functions related to *[Math Processing Error]*. Specifically, we consider functions related to the function *[Math Processing Error]* and we use the fact that

[Math Processing Error]

for *[Math Processing Error]*

✔ Example *[Math Processing Error]*: Constructing Power Series from Known Power Series

Use the power series representation for *[Math Processing Error]* combined with Theorem *[Math Processing Error]* to construct a power series for each of the following functions. Find the interval of convergence of the power series.

a. *[Math Processing Error]*
b. *[Math Processing Error]*

Solution

a. First write *[Math Processing Error]* as

[Math Processing Error]

Using the power series representation for [Math Processing Error] and parts ii. and iii. of Theorem [Math Processing Error], we find that a power series representation for [Math Processing Error] is given by

[Math Processing Error]

Since the interval of convergence of the series for [Math Processing Error] is [Math Processing Error], the interval of convergence for this new series is the set of real numbers [Math Processing Error] such that [Math Processing Error]. Therefore, the interval of convergence is [Math Processing Error]

b. To find the power series representation, use partial fractions to write [Math Processing Error] as the sum of two fractions. We have

[Math Processing Error]

First, using part ii. of Theorem [Math Processing Error], we obtain

[Math Processing Error]

Then, using parts ii. and iii. of Theorem [Math Processing Error], we have

[Math Processing Error]

Since we are combining these two power series, the interval of convergence of the difference must be the smaller of these two intervals. Using this fact and part i. of Theorem [Math Processing Error], we have

[Math Processing Error]

where the interval of convergence is [Math Processing Error].

? Exercise [Math Processing Error]

Use the series for [Math Processing Error] on [Math Processing Error] to construct a series for [Math Processing Error] Determine the interval of convergence.

Hint

Use partial fractions to rewrite [Math Processing Error] as the difference of two fractions.

Answer

[Math Processing Error]. The interval of convergence is [Math Processing Error].

In Example [Math Processing Error], we showed how to find power series for certain functions. In Example [Math Processing Error] we show how to do the opposite: given a power series, determine which function it represents.

✔ Example [Math Processing Error]: Finding the Function Represented by a Given Power Series

Consider the power series [Math Processing Error] Find the function [Math Processing Error] represented by this series. Determine the interval of convergence of the series.

Solution

Writing the given series as

[Math Processing Error]

we can recognize this series as the power series for

[Math Processing Error]

Since this is a geometric series, the series converges if and only if [Math Processing Error] Therefore, the interval of convergence is [Math Processing Error]

Exercise *[Math Processing Error]*

Find the function represented by the power series *[Math Processing Error]*.

Determine its interval of convergence.

Hint

Write *[Math Processing Error]*.

Answer

[Math Processing Error] The interval of convergence is *[Math Processing Error]*.

Recall the questions posed in the chapter opener about which is the better way of receiving payouts from lottery winnings. We now revisit those questions and show how to use series to compare values of payments over time with a lump sum payment today. We will compute how much future payments are worth in terms of today's dollars, assuming we have the ability to invest winnings and earn interest. The value of future payments in terms of today's dollars is known as the *present value* of those payments.

Example *[Math Processing Error]*: Present Value of Future Winnings

Suppose you win the lottery and are given the following three options:

- Receive 20 million dollars today;
- Receive 1.5 million dollars per year over the next 20 years; or
- Receive 1 million dollars per year indefinitely (being passed on to your heirs).

Which is the best deal, assuming that the annual interest rate is 5%? We answer this by working through the following sequence of questions.

a. How much is the 1.5 million dollars received annually over the course of 20 years worth in terms of today's dollars, assuming an annual interest rate of 5%?
b. Use the answer to part a. to find a general formula for the **present value** of payments of *[Math Processing Error]* dollars received each year over the next n years, assuming an average annual interest rate *[Math Processing Error]*.
c. Find a formula for the present value if annual payments of *[Math Processing Error]* dollars continue indefinitely, assuming an average annual interest rate *[Math Processing Error]*.
d. Use the answer to part c. to determine the present value of 1 million dollars paid annually indefinitely.
e. Use your answers to parts a. and d. to determine which of the three options is best.



Figure *[Math Processing Error]*: (credit: modification of work by Robert Huffstutter, Flickr)

Solution

a. Consider the payment of 1.5 million dollars made at the end of the first year. If you were able to receive that payment today instead of one year from now, you could invest that money and earn 5% interest. Therefore, the present value of that money *[Math Processing Error]* satisfies *[Math Processing Error]* **million dollars.** We conclude that

[Math Processing Error] **million dollars.**

Similarly, consider the payment of 1.5 million dollars made at the end of the second year. If you were able to receive that payment today, you could invest that money for two years, earning 5% interest, compounded annually. Therefore, the present value of that money *[Math Processing Error]* satisfies *[Math Processing Error]* **million dollars**. We conclude that

[Math Processing Error] **million dollars.**

The value of the future payments today is the sum of the present values *[Math Processing Error]* of each of those annual payments. The present value *[Math Processing Error]* satisfies

[Math Processing Error].

Therefore,

[Math Processing Error] **million dollars.**

b. Using the result from part a. we see that the present value *[Math Processing Error]* of *[Math Processing Error]* dollars paid annually over the course of *[Math Processing Error]* years, assuming an annual interest rate *[Math Processing Error]*, is given by

[Math Processing Error] **dollars.**

c. Using the result from part b. we see that the present value of an annuity that continues indefinitely is given by the infinite series

[Math Processing Error]

We can view the present value as a power series in *[Math Processing Error]*, which converges as long as *[Math Processing Error]*. Since *[Math Processing Error]*, this series converges. Rewriting the series as

[Math Processing Error]

we recognize this series as the power series for

[Math Processing Error].

We conclude that the present value of this annuity is

[Math Processing Error]

d. From the result to part c. we conclude that the present value *[Math Processing Error]* of *[Math Processing Error]* **million dollars** paid out every year indefinitely, assuming an annual interest rate *[Math Processing Error]*, is given by

[Math Processing Error] **million dollars.**

e. From part a. we see that receiving $1.5 million dollars over the course of 20 years is worth $18.693 million dollars in today's dollars. From part d. we see that receiving $1 million dollars per year indefinitely is worth $20 million dollars in today's dollars. Therefore, either receiving a lump-sum payment of $20 million dollars today or receiving $1 million dollars indefinitely have the same present value.

Multiplication of Power Series

We can also create new power series by multiplying power series. Being able to multiply two power series provides another way of finding power series representations for functions. The way we multiply them is similar to how we multiply polynomials. For example, suppose we want to multiply

[Math Processing Error]

and

[Math Processing Error]

It appears that the product should satisfy

[Math Processing Error]

In Theorem *[Math Processing Error]*, we state the main result regarding multiplying power series, showing that if *[Math Processing Error]* and *[Math Processing Error]* converge on a common interval *[Math Processing Error]*, then we can multiply the series in this way, and the resulting series also converges on the interval *[Math Processing Error]*.

Theorem [Math Processing Error]: Multiplying Power Series

Suppose that the power series [Math Processing Error] and [Math Processing Error] converge to [Math Processing Error] and [Math Processing Error], respectively, on a common interval [Math Processing Error]. Let

[Math Processing Error]

Then

[Math Processing Error]

and

[Math Processing Error]

The series [Math Processing Error] is known as the *Cauchy product* of the series [Math Processing Error] and [Math Processing Error].

We omit the proof of this theorem, as it is beyond the level of this text and is typically covered in a more advanced course. We now provide an example of this theorem by finding the power series representation for

[Math Processing Error]

using the power series representations for

[Math Processing Error].

Example [Math Processing Error]: Multiplying Power Series

Multiply the power series representation

[Math Processing Error]

for [Math Processing Error] with the power series representation

[Math Processing Error]

for [Math Processing Error] to construct a power series for [Math Processing Error] on the interval [Math Processing Error].

Solution

We need to multiply

[Math Processing Error]

Writing out the first several terms, we see that the product is given by

[Math Processing Error]

Since the series for [Math Processing Error] and [Math Processing Error] both converge on the interval [Math Processing Error], the series for the product also converges on the interval [Math Processing Error].

Exercise [Math Processing Error]

Multiply the series [Math Processing Error] by itself to construct a series for [Math Processing Error]

Hint

Multiply the first few terms of [Math Processing Error]

Answer

[Math Processing Error]

Differentiating and Integrating Power Series

Consider a power series *[Math Processing Error]* that converges on some interval *[Math Processing Error]*, and let *[Math Processing Error]* be the function defined by this series. Here we address two questions about *[Math Processing Error]*.

- Is *[Math Processing Error]* differentiable, and if so, how do we determine the derivative *[Math Processing Error]*?
- How do we evaluate the indefinite integral *[Math Processing Error]*?

We know that, for a polynomial with a finite number of terms, we can evaluate the derivative by differentiating each term separately. Similarly, we can evaluate the indefinite integral by integrating each term separately. Here we show that we can do the same thing for convergent power series. That is, if

[Math Processing Error]

converges on some interval *[Math Processing Error]*, then

[Math Processing Error]

and

[Math Processing Error]

Evaluating the derivative and indefinite integral in this way is called **term-by-term differentiation of a power series** and **term-by-term integration of a power series**, respectively. The ability to differentiate and integrate power series term-by-term also allows us to use known power series representations to find power series representations for other functions. For example, given the power series for *[Math Processing Error]*, we can differentiate term-by-term to find the power series for *[Math Processing Error]*. Similarly, using the power series for *[Math Processing Error]*, we can integrate term-by-term to find the power series for *[Math Processing Error]*, an antiderivative of *[Math Processing Error]*. We show how to do this in Example *[Math Processing Error]* and Example *[Math Processing Error]*. First, we state Theorem *[Math Processing Error]*, which provides the main result regarding differentiation and integration of power series.

Theorem *[Math Processing Error]*: Term-by-Term Differentiation and Integration for Power Series

Suppose that the power series *[Math Processing Error]* converges on the interval *[Math Processing Error]* for some *[Math Processing Error]*. Let *[Math Processing Error]* be the function defined by the series

[Math Processing Error]

for *[Math Processing Error]*. Then *[Math Processing Error]* is differentiable on the interval *[Math Processing Error]* and we can find *[Math Processing Error]* by differentiating the series term-by-term:

[Math Processing Error]

for *[Math Processing Error]* Also, to find *[Math Processing Error]*, we can integrate the series term-by-term. The resulting series converges on *[Math Processing Error]* and we have

[Math Processing Error]

for *[Math Processing Error]*

The proof of this result is beyond the scope of the text and is omitted. Note that although Theorem *[Math Processing Error]* guarantees the same radius of convergence when a power series is differentiated or integrated term-by-term, it says nothing about what happens at the endpoints. It is possible that the differentiated and integrated power series have different behavior at the endpoints than does the original series. We see this behavior in the next examples.

Example *[Math Processing Error]*: Differentiating Power Series

a. Use the power series representation *[Math Processing Error]* for *[Math Processing Error]* to find a power series representation for *[Math Processing Error]* on the interval *[Math Processing Error]* Determine whether the resulting series converges at the endpoints.
b. Use the result of part a. to evaluate the sum of the series *[Math Processing Error]*.

Solution

a. Since *[Math Processing Error]* is the derivative of *[Math Processing Error]*, we can find a power series representation for *[Math Processing Error]* by differentiating the power series for *[Math Processing Error]* term-by-term. The result is

[Math Processing Error]

for *[Math Processing Error]*

Theorem *[Math Processing Error]* does not guarantee anything about the behavior of this series at the endpoints. Testing the endpoints by using the divergence test, we find that the series diverges at both endpoints *[Math Processing Error]*. Note that this is the same result found in Example *[Math Processing Error]*.

b. From part a. we know that

[Math Processing Error]

Therefore,

[Math Processing Error]

? Exercise *[Math Processing Error]*

Differentiate the series *[Math Processing Error]* term-by-term to find a power series representation for *[Math Processing Error]* on the interval *[Math Processing Error]*.

Hint

Write out the first several terms and apply the power rule.

Answer

[Math Processing Error]

✓ Example *[Math Processing Error]*: Integrating Power Series

For each of the following functions *[Math Processing Error]*, find a power series representation for *[Math Processing Error]* by integrating the power series for *[Math Processing Error]* and find its interval of convergence.

a. *[Math Processing Error]*
b. *[Math Processing Error]*

Solution:

a. For *[Math Processing Error]*,the derivative is *[Math Processing Error]*. We know that

[Math Processing Error]

for *[Math Processing Error]*. To find a power series for *[Math Processing Error]*, we integrate the series term-by-term.

[Math Processing Error]

Since *[Math Processing Error]* is an antiderivative of *[Math Processing Error]*, it remains to solve for the constant *[Math Processing Error]*. Since *[Math Processing Error]*, we have *[Math Processing Error]*. Therefore, a power series representation for *[Math Processing Error]* is

[Math Processing Error]

Theorem *[Math Processing Error]* does not guarantee anything about the behavior of this power series at the endpoints. However, checking the endpoints, we find that at *[Math Processing Error]* the series is the alternating harmonic series, which converges. Also, at *[Math Processing Error]*, the series is the harmonic series, which diverges. It is important to note that, even though this series converges at *[Math Processing Error]*, Theorem *[Math Processing Error]* does not guarantee that the series actually converges to *[Math Processing Error]*. In fact, the series does converge to *[Math Processing Error]*, but showing this fact requires more advanced techniques. (Abel's theorem, covered in more advanced texts, deals with this more technical point.) The interval of convergence is *[Math Processing Error]*.

b. The derivative of *[Math Processing Error]* is *[Math Processing Error]*. We know that

[Math Processing Error]

for *[Math Processing Error]*. To find a power series for *[Math Processing Error]*, we integrate this series term-by-term.

[Math Processing Error]

Since *[Math Processing Error]*, we have *[Math Processing Error]*. Therefore, a power series representation for *[Math Processing Error]* is

[Math Processing Error]

for *[Math Processing Error]*. Again, Theorem *[Math Processing Error]* does not guarantee anything about the convergence of this series at the endpoints. However, checking the endpoints and using the alternating series test, we find that the series converges at *[Math Processing Error]* and *[Math Processing Error]*. As discussed in part a., using Abel's theorem, it can be shown that the series actually converges to *[Math Processing Error]* and *[Math Processing Error]* at *[Math Processing Error]* and *[Math Processing Error]*, respectively. Thus, the interval of convergence is *[Math Processing Error]*.

? Exercise *[Math Processing Error]*

Integrate the power series *[Math Processing Error]* term-by-term to evaluate *[Math Processing Error]*

Hint

Use the fact that *[Math Processing Error]* is an antiderivative of *[Math Processing Error]*.

Answer

[Math Processing Error]

Up to this point, we have shown several techniques for finding power series representations for functions. However, how do we know that these power series are unique? That is, given a function *[Math Processing Error]* and a power series for *[Math Processing Error]* at *[Math Processing Error]*, is it possible that there is a different power series for *[Math Processing Error]* at *[Math Processing Error]* that we could have found if we had used a different technique? The answer to this question is no. This fact should not seem surprising if we think of power series as polynomials with an infinite number of terms. Intuitively, if

[Math Processing Error]

for all values *[Math Processing Error]* in some open interval *[Math Processing Error]* about zero, then the coefficients *[Math Processing Error]* should equal *[Math Processing Error]* for *[Math Processing Error]*. We now state this result formally.

Theorem *[Math Processing Error]*: Uniqueness of Power Series

Let *[Math Processing Error]* and *[Math Processing Error]* be two convergent power series such that

[Math Processing Error]

for all *[Math Processing Error]* in an open interval containing *[Math Processing Error]*. Then *[Math Processing Error]* for all *[Math Processing Error]*.

Proof

Let

[Math Processing Error]

Then *[Math Processing Error]* By Theorem *[Math Processing Error]*, we can differentiate both series term-by-term. Therefore,

[Math Processing Error]

and thus, *[Math Processing Error]* Similarly,

[Math Processing Error]

implies that *[Math Processing Error]* and therefore, *[Math Processing Error]*. More generally, for any integer *[Math Processing Error]* and consequently, *[Math Processing Error]* for all *[Math Processing Error]*

□

In this section we have shown how to find power series representations for certain functions using various algebraic operations, differentiation, or integration. At this point, however, we are still limited as to the functions for which we can find power series representations. Next, we show how to find power series representations for many more functions by introducing Taylor series.

Key Concepts

- Given two power series *[Math Processing Error]* and *[Math Processing Error]* that converge to functions *[Math Processing Error]* and *[Math Processing Error]* on a common interval *[Math Processing Error]*, the sum and difference of the two series converge to *[Math Processing Error]*, respectively, on *[Math Processing Error]*. In addition, for any real number *[Math Processing Error]* and integer *[Math Processing Error]*, the series *[Math Processing Error]* converges to *[Math Processing Error]* and the series *[Math Processing Error]* converges to *[Math Processing Error]* whenever *[Math Processing Error]* is in the interval *[Math Processing Error]*.
- Given two power series that converge on an interval *[Math Processing Error]* the Cauchy product of the two power series converges on the interval *[Math Processing Error]*.
- Given a power series that converges to a function *[Math Processing Error]* on an interval *[Math Processing Error]*, the series can be differentiated term-by-term and the resulting series converges to *[Math Processing Error]* on *[Math Processing Error]*. The series can also be integrated term-by-term and the resulting series converges to *[Math Processing Error]* on *[Math Processing Error]*.

Glossary

term-by-term differentiation of a power series

a technique for evaluating the derivative of a power series *[Math Processing Error]* by evaluating the derivative of each term separately to create the new power series *[Math Processing Error]*

term-by-term integration of a power series

a technique for integrating a power series *[Math Processing Error]* by integrating each term separately to create the new power series *[Math Processing Error]*

This page titled 6.2: Properties of Power Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Taylor and Maclaurin Series

Learning Objectives

- Describe the procedure for finding a Taylor polynomial of a given order for a function.
- Explain the meaning and significance of Taylor's theorem with remainder.
- Estimate the remainder for a Taylor series approximation of a given function.

In the previous two sections we discussed how to find power series representations for certain types of functions—specifically, functions related to geometric series. Here we discuss power series representations for other types of functions. In particular, we address the following questions: Which functions can be represented by power series and how do we find such representations? If we can find a power series representation for a particular function f and the series converges on some interval, how do we prove that the series actually converges to f?

Overview of Taylor/Maclaurin Series

Consider a function f that has a power series representation at $x=a$. Then the series has the form

$$\sum_{n=0}^{\infty} c_n(x-a)^n = c_0 + c_1(x-a) + c_2(x-a)^2 + \ldots. \tag{6.3.1}$$

What should the coefficients be? For now, we ignore issues of convergence, but instead focus on what the series should be, if one exists. We return to discuss convergence later in this section. If the series Equation 6.3.1 is a representation for f at $x=a$, we certainly want the series to equal $f(a)$ at $x=a$. Evaluating the series at $x=a$, we see that

$$\sum_{n=0}^{\infty} c_n(x-a)^n = c_0 + c_1(a-a) + c_2(a-a)^2 + \cdots = c_0. \tag{6.3.2}$$

Thus, the series equals $f(a)$ if the coefficient $c_0 = f(a)$. In addition, we would like the first derivative of the power series to equal $f\,'(a)$ at $x=a$. Differentiating Equation 6.3.2 term-by-term, we see that

$$\frac{d}{dx}\left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \ldots. \tag{6.3.3}$$

Therefore, at $x=a$, the derivative is

$$\frac{d}{dx}\left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) = c_1 + 2c_2(a-a) + 3c_3(a-a)^2 + \cdots = c_1. \tag{6.3.4}$$

Therefore, the derivative of the series equals $f\,'(a)$ if the coefficient $c_1 = f\,'(a)$. Continuing in this way, we look for coefficients c_n such that all the derivatives of the power series Equation 6.3.4 will agree with all the corresponding derivatives of f at $x=a$. The second and third derivatives of Equation 6.3.3 are given by

$$\frac{d^2}{dx^2}\left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) = 2c_2 + 3\cdot 2c_3(x-a) + 4\cdot 3c_4(x-a)^2 + \ldots \tag{6.3.5}$$

and

$$\frac{d^3}{dx^3}\left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) = 3\cdot 2c_3 + 4\cdot 3\cdot 2c_4(x-a) + 5\cdot 4\cdot 3c_5(x-a)^2 + \cdots. \tag{6.3.6}$$

Therefore, at $x=a$, the second and third derivatives

$$\frac{d^2}{dx^2}\left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) = 2c_2 + 3\cdot 2c_3(a-a) + 4\cdot 3c_4(a-a)^2 + \cdots = 2c_2 \tag{6.3.7}$$

and

$$\frac{d^3}{dx^3}\left(\sum_{n=0}^{\infty} c_n(x-a)^n\right) = 3\cdot 2c_3 + 4\cdot 3\cdot 2c_4(a-a) + 5\cdot 4\cdot 3c_5(a-a)^2 + \cdots = 3\cdot 2c_3 \tag{6.3.8}$$

equal $f''(a)$ and $f'''(a)$, respectively, if $c_2 = \dfrac{f''(a)}{2}$ and $c_3 = \dfrac{f'''(a)}{3\cdot 2}$. More generally, we see that if f has a power series representation at $x=a$, then the coefficients should be given by $c_n = \dfrac{f^{(n)}(a)}{n!}$. That is, the series should be

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots \tag{6.3.9}$$

This power series for f is known as the Taylor series for f at a. If $x = 0$, then this series is known as the Maclaurin series for f.

Definition 6.3.1: Maclaurin and Taylor series

If f has derivatives of all orders at $x = a$, then the *Taylor series* for the function f at a is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots \tag{6.3.10}$$

The Taylor series for f at 0 is known as the *Maclaurin series* for f.

Later in this section, we will show examples of finding Taylor series and discuss conditions under which the Taylor series for a function will converge to that function. Here, we state an important result. Recall that power series representations are unique. Therefore, if a function f has a power series at a, then it must be the Taylor series for f at a.

Uniqueness of Taylor Series

If a function f has a power series at a that converges to f on some open interval containing a, then that power series is the Taylor series for f at a.

The proof follows directly from that discussed previously.

To determine if a Taylor series converges, we need to look at its sequence of partial sums. These partial sums are finite polynomials, known as **Taylor polynomials.**

Taylor Polynomials

The nth partial sum of the Taylor series for a function f at a is known as the nth Taylor polynomial. For example, the 0th, 1st, 2nd, and 3rd partial sums of the Taylor series are given by

$$p_0(x) = f(a)$$
$$p_1(x) = f(a) + f'(a)(x-a)$$
$$p_2(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2$$
$$p_3(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3$$

respectively. These partial sums are known as the 0th, 1st, 2nd, and 3rd Taylor polynomials of f at a, respectively. If $x = a$, then these polynomials are known as *Maclaurin polynomials* for f. We now provide a formal definition of Taylor and Maclaurin polynomials for a function f.

Definition 6.3.2: Maclaurin polynomial

If f has n derivatives at $x = a$, then the nth Taylor polynomial for f at a is

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{6.3.11}$$

The nth Taylor polynomial for f at 0 is known as the nth Maclaurin polynomial for f.

We now show how to use this definition to find several Taylor polynomials for $f(x) = \ln x$ at $x = 1$.

Example 6.3.1: Finding Taylor Polynomials

Find the Taylor polynomials p_0, p_1, p_2 and p_3 for $f(x) = \ln x$ at $x = 1$. Use a graphing utility to compare the graph of f with the graphs of p_0, p_1, p_2 and p_3.

Solution

To find these Taylor polynomials, we need to evaluate f and its first three derivatives at $x = 1$.

$$f(x) = \ln x \; f(1) = 0$$
$$f'(x) = \frac{1}{x} \; f'(1) = 1$$

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

$$f'''(x) = \frac{2}{x^3} \quad f'''(1) = 2$$

Therefore,

$$p_0(x) = f(1) = 0,$$

$$p_1(x) = f(1) + f'(1)(x-1) = x - 1,$$

$$p_2(x) = f(1) + f'(1)(x-1) + \frac{f''(1)}{2}(x-1)^2 = (x-1) - \frac{1}{2}(x-1)^2$$

$$p_3(x) = f(1) + f'(1)(x-1) + \frac{f''(1)}{2}(x-1)^2 + \frac{f'''(1)}{3!}(x-1)^3 = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3$$

The graphs of $y = f(x)$ and the first three Taylor polynomials are shown in Figure 6.3.1.

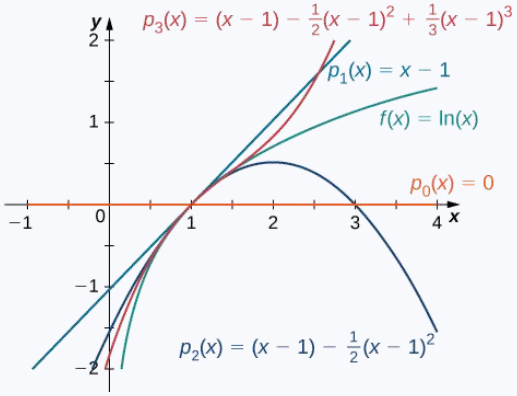


Figure 6.3.1: The function $y = \ln x$ and the Taylor polynomials p_0, p_1, p_2 and p_3 at $x = 1$ are plotted on this graph.

Exercise 6.3.1

Find the Taylor polynomials p_0, p_1, p_2 and p_3 for $f(x) = \dfrac{1}{x^2}$ at $x = 1$.

Hint

Find the first three derivatives of f and evaluate them at $x = 1$.

Answer

$$p_0(x) = 1 \tag{6.3.12}$$

$$p_1(x) = 1 - 2(x-1) \tag{6.3.13}$$

$$p_2(x) = 1 - 2(x-1) + 3(x-1)^2 \tag{6.3.14}$$

$$p_3(x) = 1 - 2(x-1) + 3(x-1)^2 - 4(x-1)^3 \tag{6.3.15}$$

We now show how to find Maclaurin polynomials for $e^x, \sin x$, and $\cos x$. As stated above, Maclaurin polynomials are Taylor polynomials centered at zero.

Example 6.3.2: Finding Maclaurin Polynomials

For each of the following functions, find formulas for the Maclaurin polynomials p_0, p_1, p_2 and p_3. Find a formula for the nth Maclaurin polynomial and write it using sigma notation. Use a graphing utility to compare the graphs of p_0, p_1, p_2 and p_3 with f.

a. $f(x) = e^x$
b. $f(x) = \sin x$
c. $f(x) = \cos x$

Solution

Since $f(x) = e^x$,we know that $f(x) = f'(x) = f''(x) = \cdots = f^{(n)}(x) = e^x$ for all positive integers n. Therefore,

$$f(0) = f'(0) = f''(0) = \cdots = f^{(n)}(0) = 1$$

for all positive integers n. Therefore, we have

$$p_0(x) = f(0) = 1,$$
$$p_1(x) = f(0) + f'(0)x = 1 + x,$$
$$p_2(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 = 1 + x + \frac{1}{2}x^2 \quad ,$$
$$p_3(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(0)}{3!}x^3 = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 \quad ,$$
$$p_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} = \sum_{k=0}^{n} \frac{x^k}{k!} \quad .$$

The function and the first three Maclaurin polynomials are shown in Figure 2.

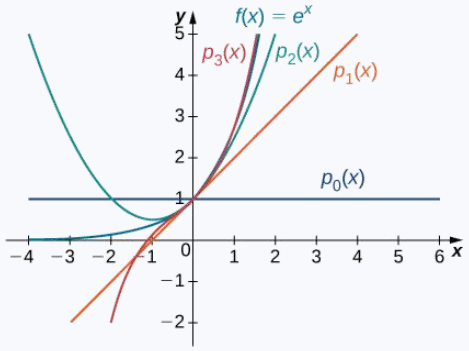


Figure 6.3.2: *The graph shows the function $y = e^x$ and the Maclaurin polynomials p_0, p_1, p_2 and p_3.*

b. For $f(x) = \sin x$, the values of the function and its first four derivatives at $x = 0$ are given as follows:

$$f(x) = \sin x \ f(0) = 0$$
$$f'(x) = \cos x \ f'(0) = 1$$
$$f''(x) = -\sin x \ f''(0) = 0$$
$$f'''(x) = -\cos x \ f'''(0) = -1$$
$$f^{(4)}(x) = \sin x \ f^{(4)}(0) = 0.$$

Since the fourth derivative is $\sin x$, the pattern repeats. That is, $f^{(2m)}(0) = 0$ and $f^{(2m+1)}(0) = (-1)^m$ for $m \geq 0$. Thus, we have

$$p_0(x) = 0,$$
$$p_1(x) = 0 + x = x,$$
$$p_2(x) = 0 + x + 0 = x,$$
$$p_3(x) = 0 + x + 0 - \frac{1}{3!}x^3 = x - \frac{x^3}{3!},$$
$$p_4(x) = 0 + x + 0 - \frac{1}{3!}x^3 + 0 = x - \frac{x^3}{3!} \quad ,$$
$$p_5(x) = 0 + x + 0 - \frac{1}{3!}x^3 + 0 + \frac{1}{5!}x^5 = x - \frac{x^3}{3!} + \frac{x^5}{5!} \quad ,$$

and for $m \geq 0$,

$$p_{2m+1}(x) = p_{2m+2}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^m \frac{x^{2m+1}}{(2m+1)!} = \sum_{k=0}^{m} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \quad .$$

Graphs of the function and its Maclaurin polynomials are shown in Figure 3.

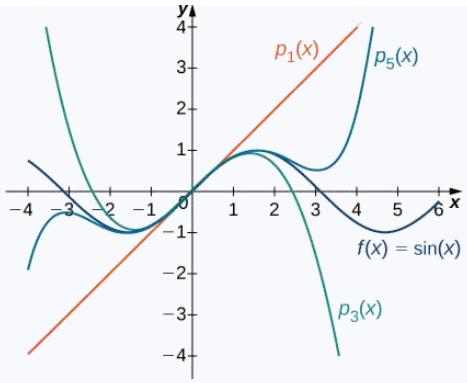


Figure 6.3.3: *The graph shows the function $y = \sin x$ and the Maclaurin polynomials p_1, p_3 and p_5.*

c. For $f(x) = \cos x$, the values of the function and its first four derivatives at $x = 0$ are given as follows:

$$f(x) = \cos x \; f(0) = 1$$
$$f'(x) = -\sin x \; f'(0) = 0$$
$$f''(x) = -\cos x \; f''(0) = -1$$
$$f'''(x) = \sin x \; f'''(0) = 0$$
$$f^{(4)}(x) = \cos x \; f^{(4)}(0) = 1.$$

Since the fourth derivative is $\sin x$, the pattern repeats. In other words, $f^{(2m)}(0) = (-1)^m$ and $f^{(2m+1)} = 0$ for $m \geq 0$. Therefore,

$$p_0(x) = 1,$$
$$p_1(x) = 1 + 0 = 1,$$
$$p_2(x) = 1 + 0 - \frac{1}{2!}x^2 = 1 - \frac{x^2}{2!} \quad ,$$
$$p_3(x) = 1 + 0 - \frac{1}{2!}x^2 + 0 = 1 - \frac{x^2}{2!} \quad ,$$
$$p_4(x) = 1 + 0 - \frac{1}{2!}x^2 + 0 + \frac{1}{4!}x^4 = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \quad ,$$
$$p_5(x) = 1 + 0 - \frac{1}{2!}x^2 + 0 + \frac{1}{4!}x^4 + 0 = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \quad ,$$

and for $n \geq 0$,

$$p_{2m}(x) = p_{2m+1}(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^m \frac{x^{2m}}{(2m)!} = \sum_{k=0}^{m} (-1)^k \frac{x^{2k}}{(2k)!} \quad .$$

Graphs of the function and the Maclaurin polynomials appear in Figure 4.

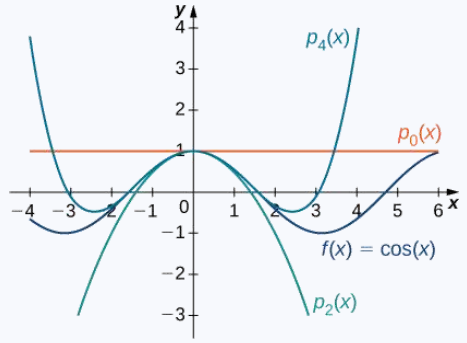


Figure 6.3.4: *The function $y = \cos x$ and the Maclaurin polynomials p_0, p_2 and p_4 are plotted on this graph.*

Exercise 6.3.2

Find formulas for the Maclaurin polynomials p_0, p_1, p_2 and p_3 for $f(x) = \dfrac{1}{1+x}$.

Find a formula for the *n*th Maclaurin polynomial. Write your answer using sigma notation.

Hint

Evaluate the first four derivatives of f and look for a pattern.

Answer

$p_0(x)=1; p_1(x)=1-x; p_2(x)=1-x+x^2; p_3(x)=1-x+x^2-x^3; p_n(x)=1-x+x^2-x^3+\cdots+(-1)^n x^n=\sum_{k=0}^{n}(-1)^k x^k$

Taylor's Theorem with Remainder

Recall that the *n*th Taylor polynomial for a function f at a is the nth partial sum of the Taylor series for f at a. Therefore, to determine if the Taylor series converges, we need to determine whether the sequence of Taylor polynomials p_n converges. However, not only do we want to know if the sequence of Taylor polynomials converges, we want to know if it converges to f. To answer this question, we define the remainder $R_n(x)$ as

$$R_n(x)=f(x)-p_n(x). \tag{6.3.16}$$

For the sequence of Taylor polynomials to converge to f, we need the remainder R_n to converge to zero. To determine if R_n converges to zero, we introduce **Taylor's theorem with remainder**. Not only is this theorem useful in proving that a Taylor series converges to its related function, but it will also allow us to quantify how well the *n*th Taylor polynomial approximates the function.

Here we look for a bound on $|R_n|$. Consider the simplest case: $n=0$. Let p_0 be the 0th Taylor polynomial at a for a function f. The remainder R_0 satisfies

$$R_0(x)=f(x)-p_0(x)=f(x)-f(a).$$

If f is differentiable on an interval *I* containing a and x, then by the Mean Value Theorem there exists a real number *c* between *a* and *x* such that $f(x)-f(a)=f'(c)(x-a)$. Therefore,

$$R_0(x)=f'(c)(x-a). \tag{6.3.17}$$

Using the Mean Value Theorem in a similar argument, we can show that if f is *n* times differentiable on an interval *I* containing *a* and *x*, then the *n*th remainder R_n satisfies

$$R_n(x)=\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{6.3.18}$$

for some real number *c* between a and *x*. It is important to note that the value *c* in the numerator above is not the center *a*, but rather an unknown value c between *a* and *x*. This formula allows us to get a bound on the remainder R_n. If we happen to know that $\left|f^{(n+1)}(x)\right|$ is bounded by some real number *M* on this interval *I*, then

$$|R_n(x)|\le\frac{M}{(n+1)!}|x-a|^{n+1} \tag{6.3.19}$$

for all *x* in the interval *I*.

We now state Taylor's theorem, which provides the formal relationship between a function f and its *n*th degree Taylor polynomial $p_n(x)$. This theorem allows us to bound the error when using a Taylor polynomial to approximate a function value, and will be important in proving that a Taylor series for f converges to f.

Taylor's Theorem with Remainder

Let f be a function that can be differentiated $n+1$ times on an interval *I* containing the real number *a*. Let p_n be the *n*th Taylor polynomial of f at *a* and let

$$R_n(x)=f(x)-p_n(x) \tag{6.3.20}$$

be the *n*th remainder. Then for each *x* in the interval *I*, there exists a real number *c* between *a* and x such that

$$R_n(x)=\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{6.3.21}$$

.

If there exists a real number M such that $\left|f^{(n+1)}(x)\right|\le M$ for all $x\in I$, then

$$|R_n(x)|\le\frac{M}{(n+1)!}|x-a|^{n+1} \tag{6.3.22}$$

for all x in I.

Proof

Fix a point $x \in I$ and introduce the function g such that

$$g(t) = f(x) - f(t) - f'(t)(x-t) - \frac{f''(t)}{2!}(x-t)^2 - \cdots - \frac{f^{(n)}(t)}{n!}(x-t)^n - R_n(x)\frac{(x-t)^{n+1}}{(x-a)^{n+1}}. \tag{6.3.23}$$

We claim that g satisfies the criteria of Rolle's theorem. Since \(g\) is a polynomial function (in t), it is a differentiable function. Also, g is zero at $t = a$ and $t = x$ because

$$g(a) = f(x) - f(a) - f'(a)(x-a) - \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n - R_n(x) \tag{6.3.24}$$

$$= f(x) - p_n(x) - R_n(x) \tag{6.3.25}$$

$$= 0, \tag{6.3.26}$$

$$g(x) = f(x) - f(x) - 0 - \cdots - 0 \tag{6.3.27}$$

$$= 0. \tag{6.3.28}$$

Therefore, g satisfies Rolle's theorem, and consequently, there exists c between a and x such that $g'(c) = 0$. We now calculate g'. Using the product rule, we note that

$$\frac{d}{dt}\left[\frac{f^{(n)}(t)}{n!}(x-t)^n\right] = -\frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} + \frac{f^{(n+1)}(t)}{n!}(x-t)^n. \tag{6.3.29}$$

Consequently,

$$g'(t) = -f'(t) + [f'(t) - f''(t)(x-t)] + [f''(t)(x-t) - \frac{f'''(t)}{2!}(x-t)^2] + \cdots \tag{6.3.30}$$

$$+ [\frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} - \frac{f^{(n+1)}(t)}{n!}(x-t)^n] + (n+1)R_n(x)\frac{(x-t)^n}{(x-a)^{n+1}}$$

.

Notice that there is a telescoping effect. Therefore,

$$g'(t) = -\frac{f^{(n+1)}(t)}{n!}(x-t)^n + (n+1)R_n(x)\frac{(x-t)^n}{(x-a)^{n+1}} \tag{6.3.31}$$

.

By Rolle's theorem, we conclude that there exists a number c between a and x such that $g'(c) = 0$. Since

$$g'(c) = -\frac{f^{(n+1)}(c)}{n!}(x-c)^n + (n+1)R_n(x)\frac{(x-c)^n}{(x-a)^{n+1}} \tag{6.3.32}$$

we conclude that

$$-\frac{f^{(n+1)}(c)}{n!}(x-c)^n + (n+1)R_n(x)\frac{(x-c)^n}{(x-a)^{n+1}} = 0. \tag{6.3.33}$$

Adding the first term on the left-hand side to both sides of the equation and dividing both sides of the equation by $n+1$, we conclude that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{6.3.34}$$

as desired. From this fact, it follows that if there exists M such that $\left|f^{(n+1)}(x)\right| \leq M$ for all x in I, then

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1} \tag{6.3.35}$$

.

□

Not only does Taylor's theorem allow us to prove that a Taylor series converges to a function, but it also allows us to estimate the accuracy of Taylor polynomials in approximating function values. We begin by looking at linear and quadratic approximations of $f(x) = \sqrt[3]{x}$ at $x = 8$ and

determine how accurate these approximations are at estimating $\sqrt[3]{11}$.

Example 6.3.3: Using Linear and Quadratic Approximations to Estimate Function Values

Consider the function $f(x) = \sqrt[3]{x}$.

a. Find the first and second Taylor polynomials for f at $x = 8$. Use a graphing utility to compare these polynomials with f near $x = 8$.
b. Use these two polynomials to estimate $\sqrt[3]{11}$.
c. Use Taylor's theorem to bound the error.

Solution:

a. For $f(x) = \sqrt[3]{x}$, the values of the function and its first two derivatives at $x = 8$ are as follows:

$$f(x) = \sqrt[3]{x} \quad f(8) = 2$$
$$f'(x) = \frac{1}{3x^{2/3}} \quad f'(8) = \frac{1}{12}$$
$$f''(x) = \frac{-2}{9x^{5/3}} \quad f''(8) = -\frac{1}{144}.$$

Thus, the first and second Taylor polynomials at $x = 8$ are given by

$$\begin{aligned} p_1(x) &= f(8) + f'(8)(x-8) \\ &= 2 + \frac{1}{12}(x-8) \\ p_2(x) &= f(8) + f'(8)(x-8) + \frac{f''(8)}{2!}(x-8)^2 \\ &= 2 + \frac{1}{12}(x-8) - \frac{1}{288}(x-8)^2 \ . \end{aligned}$$

The function and the Taylor polynomials are shown in Figure 4.

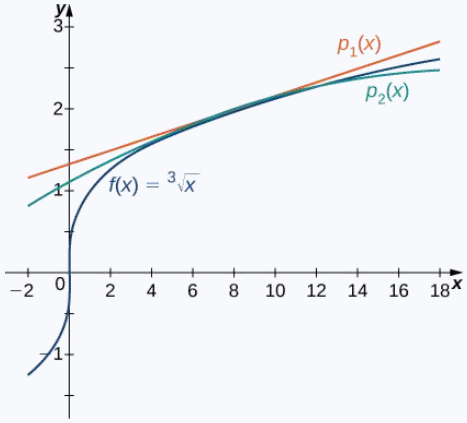


Figure 6.3.5: *The graphs of $f(x) = \sqrt[3]{x}$ and the linear and quadratic approximations $p_1(x)$ and $p_2(x)$*

b. Using the first Taylor polynomial at $x = 8$, we can estimate

$$\sqrt[3]{11} \approx p_1(11) = 2 + \frac{1}{12}(11-8) = 2.25. \tag{6.3.36}$$

Using the second Taylor polynomial at $x = 8$, we obtain

$$\sqrt[3]{11} \approx p_2(11) = 2 + \frac{1}{12}(11-8) - \frac{1}{288}(11-8)^2 = 2.21875. \tag{6.3.37}$$

c. By Note, there exists a c in the interval $(8, 11)$ such that the remainder when approximating $\sqrt[3]{11}$ by the first Taylor polynomial satisfies

$$R_1(11) = \frac{f''(c)}{2!}(11-8)^2. \tag{6.3.38}$$

We do not know the exact value of c, so we find an upper bound on $R_1(11)$ by determining the maximum value of f'' on the interval $(8, 11)$. Since $f''(x) = -\frac{2}{9x^{5/3}}$, the largest value for $|f''(x)|$ on that interval occurs at $x = 8$. Using the fact that $f''(8) = -\frac{1}{144}$, we obtain

$$|R_1(11)| \le \frac{1}{144 \cdot 2!}(11-8)^2 = 0.03125.$$

Similarly, to estimate $R_2(11)$, we use the fact that

$$R_2(11) = \frac{f'''(c)}{3!}(11-8)^3.$$

Since $f'''(x) = \dfrac{10}{27x^{8/3}}$, the maximum value of f''' on the interval $(8, 11)$ is $f'''(8) \approx 0.0014468$. Therefore, we have

$$|R_2(11)| \le \frac{0.0011468}{3!}(11-8)^3 \approx 0.0065104.$$

Exercise 6.3.3:

Find the first and second Taylor polynomials for $f(x) = \sqrt{x}$ at $x = 4$. Use these polynomials to estimate $\sqrt{6}$. Use Taylor's theorem to bound the error.

Hint

Evaluate $f(4)$, $f\,'(4)$,and $f''(4)$.

Answer

$p_1(x) = 2 + \frac{1}{4}(x-4); p_2(x) = 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2; p_1(6) = 2.5; p_2(6) = 2.4375;$
$|R_1(6)| \le 0.0625; |R_2(6)| \le 0.015625$

Example 6.3.4: Approximating $\sin x$ Using Maclaurin Polynomials

From Example b., the Maclaurin polynomials for $\sin x$ are given by

$$p_{2m+1}(x) = p_{2m+2}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^m \frac{x^{2m+1}}{(2m+1)!}$$

for $m = 0, 1, 2, \ldots$.

a. Use the fifth Maclaurin polynomial for $\sin x$ to approximate $sin(\frac{\pi}{18})$ and bound the error.

b. For what values of x does the fifth Maclaurin polynomial approximate $\sin x$ to within 0.0001?

Solution

a.

The fifth Maclaurin polynomial is

$$p_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} \tag{6.3.39}$$

.

Using this polynomial, we can estimate as follows:

$$\sin(\frac{\pi}{18}) \approx p_5(\frac{\pi}{18}) = \frac{\pi}{18} - \frac{1}{3!}(\frac{\pi}{18})^3 + \frac{1}{5!}(\frac{\pi}{18})^5 \approx 0.173648. \tag{6.3.40}$$

To estimate the error, use the fact that the sixth Maclaurin polynomial is $p_6(x) = p_5(x)$ and calculate a bound on $R_6(\frac{\pi}{18})$. By Note, the remainder is

$$R_6(\frac{\pi}{18}) = \frac{f^{(7)}(c)}{7!}(\frac{\pi}{18})^7 \tag{6.3.41}$$

for some c between 0 and $\frac{\pi}{18}$. Using the fact that $\left|f^{(7)}(x)\right| \le 1$ for all x, we find that the magnitude of the error is at most

$$\frac{1}{7!} \cdot (\frac{\pi}{18})^7 \le 9.8 \times 10^{-10}. \tag{6.3.42}$$

b.

We need to find the values of x such that

$$\frac{1}{7}!|x|^7 \leq 0.0001. \tag{6.3.43}$$

Solving this inequality for x, we have that the fifth Maclaurin polynomial gives an estimate to within 0.0001 as long as $|x| < 0.907$.

Exercise 6.3.4

Use the fourth Maclaurin polynomial for $\cos x$ to approximate $\cos(\frac{\pi}{12})$.

Hint

The fourth Maclaurin polynomial is $p_4(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$.

Answer

0.96593

Now that we are able to bound the remainder $R_n(x)$, we can use this bound to prove that a Taylor series for f at a converges to f.

Representing Functions with Taylor and Maclaurin Series

We now discuss issues of convergence for Taylor series. We begin by showing how to find a Taylor series for a function, and how to find its interval of convergence.

Example 6.3.5: Finding a Taylor Series

Find the Taylor series for $f(x) = \frac{1}{x}$ at $x = 1$. Determine the interval of convergence.

Solution

For $f(x) = \frac{1}{x}$, the values of the function and its first four derivatives at $x = 1$ are

$$f(x) = \frac{1}{x} \; f(1) = 1$$

$$f\,'(x) = -\frac{1}{x^2} \; f\,'(1) = -1$$

$$f''(x) = \frac{2}{x^3} \; f''(1) = 2!$$

$$f'''(x) = -\frac{3 \cdot 2}{x^4} \; f'''(1) = -3!$$

$$f^{(4)}(x) = \frac{4 \cdot 3 \cdot 2}{x^5} \; f^{(4)}(1) = 4!.$$

That is, we have $f^{(n)}(1) = (-1)^n n!$ for all $n \geq 0$. Therefore, the Taylor series for f at $x = 1$ is given by

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(1)}{n!}(x-1)^n = \sum_{n=0}^{\infty}(-1)^n(x-1)^n \; .$$

To find the interval of convergence, we use the ratio test. We find that

$$\frac{|a_{n+1}|}{|a_n|} = \frac{\left|(-1)^{n+1}(x-1)n^{+1}\right|}{|(-1)^n(x-1)^n|} = |x-1| \; .$$

Thus, the series converges if $|x-1| < 1$. That is, the series converges for $0 < x < 2$. Next, we need to check the endpoints. At $x = 2$, we see that

$$\sum_{n=0}^{\infty}(-1)^n(2-1)^n = \sum_{n=0}^{\infty}(-1)^n$$

diverges by the divergence test. Similarly, at $x = 0$,

$$\sum_{n=0}^{\infty}(-1)^n(0-1)^n = \sum_{n=0}^{\infty}(-1)^{2n} = \sum_{n=0}^{\infty}1$$

diverges. Therefore, the interval of convergence is $(0, 2)$.

Exercise 6.3.5

Find the Taylor series for $f(x)=\dfrac{1}{2}$ at $x=2$ and determine its interval of convergence.

Hint

$f^{(n)}(2)=\dfrac{(-1)^n n!}{2^{n+1}}$

Answer

$\dfrac{1}{2}\displaystyle\sum_{n=0}^{\infty}(\dfrac{2-x}{2})^n$. The interval of convergence is $(0,4)$.

We know that the Taylor series found in this example converges on the interval $(0,2)$, but how do we know it actually converges to f? We consider this question in more generality in a moment, but for this example, we can answer this question by writing

$$f(x)=\frac{1}{x}=\frac{1}{1-(1-x)}. \tag{6.3.44}$$

That is, f can be represented by the geometric series $\displaystyle\sum_{n=0}^{\infty}(1-x)^n$. Since this is a geometric series, it converges to $\dfrac{1}{x}$ as long as $|1-x|<1$. Therefore, the Taylor series found in Example does converge to $f(x)=\dfrac{1}{x}$ on $(0,2)$.

We now consider the more general question: if a Taylor series for a function f converges on some interval, how can we determine if it actually converges to f? To answer this question, recall that a series converges to a particular value if and only if its sequence of partial sums converges to that value. Given a Taylor series for f at a, the nth partial sum is given by the nth Taylor polynomial pn. Therefore, to determine if the Taylor series converges to f, we need to determine whether

$$\lim_{n\to\infty}p_n(x)=f(x).$$

Since the remainder $R_n(x)=f(x)-p_n(x)$, the Taylor series converges to f if and only if

$$\lim_{n\to\infty}R_n(x)=0.$$

We now state this theorem formally.

Convergence of Taylor Series

Suppose that f has derivatives of all orders on an interval I containing a. Then the Taylor series

$$\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n \tag{6.3.45}$$

converges to $f(x)$ for all x in I if and only if

$$\lim_{n\to\infty}R_n(x)=0 \tag{6.3.46}$$

for all x in I.

With this theorem, we can prove that a Taylor series for f at a converges to f if we can prove that the remainder $R_n(x)\to 0$. To prove that $R_n(x)\to 0$, we typically use the bound

$$|R_n(x)|\le\frac{M}{(n+1)!}|x-a|^{n+1} \tag{6.3.47}$$

from Taylor's theorem with remainder.

In the next example, we find the Maclaurin series for e^x and $\sin x$ and show that these series converge to the corresponding functions for all real numbers by proving that the remainders $R_n(x)\to 0$ for all real numbers x.

Example 6.3.6: Finding Maclaurin Series

For each of the following functions, find the Maclaurin series and its interval of convergence. Use Note to prove that the Maclaurin series for f converges to f on that interval.

a. e^x
b. $\sin x$

Solution

a. Using the *n*th Maclaurin polynomial for e^x found in Example a., we find that the Maclaurin series for e^x is given by

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

To determine the interval of convergence, we use the ratio test. Since

$$\frac{|a_{n+1}|}{|a_n|} = \frac{|x|^{n+1}}{(n+1)!} \cdot \frac{n!}{|x|^n} = \frac{|x|}{n+1},$$

we have

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \frac{|x|}{n+1} = 0$$

for all x. Therefore, the series converges absolutely for all x, and thus, the interval of convergence is $(-\infty, \infty)$. To show that the series converges to e^x for all x, we use the fact that $f^{(n)}(x) = e^x$ for all $n \geq 0$ and e^x is an increasing function on $(-\infty, \infty)$. Therefore, for any real number b, the maximum value of e^x for all $|x| \leq b$ is e^b. Thus,

$$|R_n(x)| \leq \frac{e^b}{(n+1)!}|x|^{n+1}.$$

Since we just showed that

$$\sum_{n=0}^{\infty} \frac{|x|^n}{n!}$$

converges for all *x*, by the divergence test, we know that

$$\lim_{n\to\infty} \frac{|x|^{n+1}}{(n+1)!} = 0$$

for any real number *x*. By combining this fact with the squeeze theorem, the result is $\lim_{n\to\infty} R_n(x) = 0$.

b. Using the *n*th Maclaurin polynomial for $\sin x$ found in Example b., we find that the Maclaurin series for $\sin x$ is given by

$$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

In order to apply the ratio test, consider

$$\frac{|a_{n+1}|}{|a_n|} = \frac{|x|^{2n+3}}{(2n+3)!} \cdot \frac{(2n+1)!}{|x|^{2n+1}} = \frac{|x|^2}{(2n+3)(2n+2)}.$$

Since

$$\lim_{n\to\infty} \frac{|x|^2}{(2n+3)(2n+2)} = 0$$

for all x, we obtain the interval of convergence as $(-\infty, \infty)$. To show that the Maclaurin series converges to $\sin x$, look at $R_n(x)$. For each x there exists a real number c between 0 and x such that

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1}.$$

Since $\left|f^{(n+1)}(c)\right| \leq 1$ for all integers n and all real numbers c, we have

$$|R_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!}$$

for all real numbers x. Using the same idea as in part a., the result is $\lim_{n\to\infty} R_n(x) = 0$ for all x, and therefore, the Maclaurin series for $\sin x$ converges to $\sin x$ for all real x.

Exercise 6.3.6

Find the Maclaurin series for $f(x) = \cos x$. Use the ratio test to show that the interval of convergence is $(-\infty, \infty)$. Show that the Maclaurin series converges to $\cos x$ for all real numbers x.

Hint

Use the Maclaurin polynomials for $\cos x$.

Answer

$\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$

By the ratio test, the interval of convergence is $(-\infty, \infty)$. Since $|R_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!}$, the series converges to $\cos x$ for all real x.

Proving that *E* is Irrational

In this project, we use the Maclaurin polynomials for e^x to prove that e is irrational. The proof relies on supposing that e is rational and arriving at a contradiction. Therefore, in the following steps, we suppose $e = r/s$ for some integers r and s where $s \neq 0$.

1. Write the Maclaurin polynomials $p_0(x), p_1(x), p_2(x), p_3(x), p_4(x)$for e^x. Evaluate $p_0(1), p_1(1), p_2(1), p_3(1), p_4(1)$to estimate e.
2. Let $R_n(x)$ denote the remainder when using $p_n(x)$ to estimate e^x. Therefore, $R_n(x) = e^x - p_n(x)$, and $R_n(1) = e - p_n(1)$. Assuming that $e = \frac{r}{s}$ for integers r and s, evaluate $R_0(1), R_1(1), R_2(1), R_3(1), R_4(1)$.
3. Using the results from part 2, show that for each remainder $R_0(1), R_1(1), R_2(1), R_3(1), R_4(1)$, we can find an integer k such that $kR_n(1)$ is an integer for $n = 0, 1, 2, 3, 4$.
4. Write down the formula for the nth Maclaurin polynomial $p_n(x)$ for e^x and the corresponding remainder $R_n(x)$. Show that $sn!R_n(1)$ is an integer.
5. Use Taylor's theorem to write down an explicit formula for $R_n(1)$. Conclude that $R_n(1) \neq 0$, and therefore, $sn!R_n(1) \neq 0$.
6. Use Taylor's theorem to find an estimate on $R_n(1)$. Use this estimate combined with the result from part 5 to show that $|sn!R_n(1)| < \frac{se}{n+1}$. Conclude that if *n* is large enough, then $|sn!R_n(1)| < 1$. Therefore, $sn!R_n(1)$ is an integer with magnitude less than 1. Thus, $sn!R_n(1) = 0$. But from part 5, we know that $sn!R_n(1) \neq 0$. We have arrived at a contradiction, and consequently, the original supposition that e is rational must be false.

Key Concepts

- Taylor polynomials are used to approximate functions near a value $x = a$. Maclaurin polynomials are Taylor polynomials at $x = 0$.
- The *n*th degree Taylor polynomials for a function f are the partial sums of the Taylor series for f.
- If a function f has a power series representation at $x = a$, then it is given by its Taylor series at $x = a$.
- A Taylor series for f converges to f if and only if $\lim_{n\to\infty} R_n(x) = 0$ where $R_n(x) = f(x) - p_n(x)$.
- The Taylor series for e^x, $\sin x$, and $\cos x$ converge to the respective functions for all real x.

Key Equations

- **Taylor series for the function f at the point $x = a$**

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots$$

Glossary

Maclaurin polynomial

a Taylor polynomial centered at 0; the nth Taylor polynomial for f at 0 is the nth Maclaurin polynomial for f

Maclaurin series

a Taylor series for a function f at $x = 0$ is known as a Maclaurin series for f

Taylor polynomials

the *n*th Taylor polynomial for f at $x = a$ is $p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$

Taylor series

a power series at a that converges to a function f on some open interval containing a

Taylor's theorem with remainder

for a function f and the *n*th Taylor polynomial for f at $x = a$, the remainder $R_n(x) = f(x) - p_n(x)$ satisfies $R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$

for some c between x and a; if there exists an interval I containing a and a real number M such that $\left|f^{(n+1)}(x)\right| \leq M$ for all x in I, then $|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}$

Contributors

- Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 6.3: Taylor and Maclaurin Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

6.4: Working with Taylor Series

Learning Objectives

- Write the terms of the binomial series.
- Recognize the Taylor series expansions of common functions.
- Recognize and apply techniques to find the Taylor series for a function.
- Use Taylor series to solve differential equations.
- Use Taylor series to evaluate non-elementary integrals.

In the preceding section, we defined Taylor series and showed how to find the Taylor series for several common functions by explicitly calculating the coefficients of the Taylor polynomials. In this section we show how to use those Taylor series to derive Taylor series for other functions. We then present two common applications of power series. First, we show how power series can be used to solve differential equations. Second, we show how power series can be used to evaluate integrals when the antiderivative of the integrand cannot be expressed in terms of elementary functions. In one example, we consider $\int e^{-x^2}\,dx$, an integral that arises frequently in probability theory.

The Binomial Series

Our first goal in this section is to determine the Maclaurin series for the function $f(x)=(1+x)^r$ for all real numbers r. The Maclaurin series for this function is known as the **binomial series.** We begin by considering the simplest case: r is a nonnegative integer. We recall that, for $r=0,\ 1,\ 2,\ 3,\ 4,\ f(x)=(1+x)^r$ can be written as

$$\begin{aligned} f(x) &= (1+x)^0 = 1, \\ f(x) &= (1+x)^1 = 1+x, \\ f(x) &= (1+x)^2 = 1+2x+x^2, \\ f(x) &= (1+x)^3 = 1+3x+3x^2+x^3 \\ f(x) &= (1+x)^4 = 1+4x+6x^2+4x^3+x^4. \end{aligned}$$

The expressions on the right-hand side are known as binomial expansions and the coefficients are known as binomial coefficients. More generally, for any nonnegative integer r, the binomial coefficient of x^n in the binomial expansion of $(1+x)^r$ is given by

$$\binom{r}{n} = \frac{r!}{n!(r-n)!} \tag{6.4.1}$$

and

$$\begin{aligned} f(x) &= (1+x)^r \\ &= \binom{r}{0} + \binom{r}{1}x + \binom{r}{2}x^2 + \binom{r}{3}x^3 + \cdots + \binom{r}{r-1}x^{r-1} + \binom{r}{r}x^r \\ &= \sum_{n=0}^{r}\binom{r}{n}x^n. \end{aligned} \tag{6.4.2}$$

For example, using this formula for $r=5$, we see that

$$\begin{aligned} f(x) &= (1+x)^5 \\ &= \binom{5}{0}1 + \binom{5}{1}x + \binom{5}{2}x^2 + \binom{5}{3}x^3 + \binom{5}{4}x^4 + \binom{5}{5}x^5 \\ &= \frac{5!}{0!5!}1 + \frac{5!}{1!4!}x + \frac{5!}{2!3!}x^2 + \frac{5!}{3!2!}x^3 + \frac{5!}{4!1!}x^4 + \frac{5!}{5!0!}x^5 \\ &= 1+5x+10x^2+10x^3+5x^4+x^5. \end{aligned}$$

We now consider the case when the exponent r is any real number, not necessarily a nonnegative integer. If r is not a nonnegative integer, then $f(x)=(1+x)^r$ cannot be written as a finite polynomial. However, we can find a power series for f. Specifically, we look for the Maclaurin series for f. To do this, we find the derivatives of f and evaluate them at $x=0$.

$$\begin{aligned} f(x) &= (1+x)^r & f(0) &= 1 \\ f'(x) &= r(1+x)^{r-1} & f'(0) &= r \\ f''(x) &= r(r-1)(1+x)^{r-2} & f''(0) &= r(r-1) \\ f'''(x) &= r(r-1)(r-2)(1+x)^{r-3} & f'''(0) &= r(r-1)(r-2) \\ f^{(n)}(x) &= r(r-1)(r-2)\cdots(r-n+1)(1+x)^{r-n} & f^{(n)}(0) &= r(r-1)(r-2)\cdots(r-n+1) \end{aligned}$$

We conclude that the coefficients in the binomial series are given by

$$\frac{f^{(n)}(0)}{n!}=\frac{r(r-1)(r-2)\cdots(r-n+1)}{n!}. \tag{6.4.3}$$

We note that if r is a nonnegative integer, then the $(r+1)^{\text{st}}$ derivative $f^{(r+1)}$ is the zero function, and the series terminates. In addition, if r is a nonnegative integer, then Equation 6.4.3 for the coefficients agrees with Equation 6.4.1 for the coefficients, and the formula for the binomial series agrees with Equation 6.4.2 for the finite binomial expansion. More generally, to denote the binomial coefficients for any real number r, we define

$$\binom{r}{n}=\frac{(r-1)(r-2)\cdots(r-n+1)}{n!}.$$

With this notation, we can write the binomial series for $(1+x)^r$ as

$$\sum_{n=0}^{\infty}\binom{r}{n}x^n=1+rx+\frac{r(r-1)}{2!}x^2+\cdots+\frac{r(r-1)\cdots(r-n+1)}{n!}x^n+\cdots. \tag{6.4.4}$$

We now need to determine the interval of convergence for the binomial series Equation 6.4.4. We apply the ratio test. Consequently, we consider

$$\begin{aligned} \frac{|a_{n+1}|}{|a_n|} &= \frac{|r(r-1)(r-2)\cdots(r-n)||x||^{n+1}}{(n+1)!}\cdot\frac{n!}{|r(r-1)(r-2)\cdots(r-n+1)||x|^n} \\ &= \frac{|r-n||x|}{|n+1|} \end{aligned}$$

.

Since

$$\lim_{n\to\infty}\frac{|a_{n+1}|}{|a_n|}=|x|<1$$

if and only if $|x|<1$, we conclude that the interval of convergence for the binomial series is $(-1,1)$. The behavior at the endpoints depends on r. It can be shown that for $r\geq 0$ the series converges at both endpoints; for $-1<r<0$, the series converges at $x=1$ and diverges at $x=-1$; and for $r<-1$, the series diverges at both endpoints. The binomial series does converge to $(1+x)^r$ in $(-1,1)$ for all real numbers r, but proving this fact by showing that the remainder $R_n(x)\to 0$ is difficult.

Definition: Binomial Series

For any real number r, the Maclaurin series for $f(x)=(1+x)^r$ is the binomial series. It converges to f for $|x|<1$, and we write

$$(1+x)^r=\sum_{n=0}^{\infty}\binom{r}{n}x^n=1+rx+\frac{r(r-1)}{2!}x^2+\cdots+r\frac{(r-1)\cdots(r-n+1)}{n!}x^n+\cdots$$

for $|x|<1$.

We can use this definition to find the binomial series for $f(x)=\sqrt{1+x}$ and use the series to approximate $\sqrt{1.5}$.

Example 6.4.1: Finding Binomial Series

a. Find the binomial series for $f(x)=\sqrt{1+x}$.
b. Use the third-order Maclaurin polynomial $p_3(x)$ to estimate $\sqrt{1.5}$. Use Taylor's theorem to bound the error. Use a graphing utility to compare the graphs of f and p_3.

Solution

a. Here $r=\dfrac{1}{2}$. Using the definition for the binomial series, we obtain

$$\begin{aligned}\sqrt{1+x} &= 1+\frac{1}{2}x+\frac{(1/2)(-1/2)}{2!}x^2+\frac{(1/2)(-1/2)(-3/2)}{3!}x^3+\cdots \\ &= 1+\frac{1}{2}x-\frac{1}{2!}\frac{1}{2^2}x^2+\frac{1}{3!}\frac{1\cdot 3}{2^3}x^3-\cdots+\frac{(-1)^{n+1}}{n!}\frac{1\cdot 3\cdot 5\cdots(2n-3)}{2^n}x^n+\cdots \\ &= 1+\frac{1}{2}x+\sum_{n=2}^{\infty}\frac{(-1)^{n+1}}{n!}\frac{1\cdot 3\cdot 5\cdots(2n-3)}{2^n}x^n.\end{aligned}$$

b. From the result in part a. the third-order Maclaurin polynomial is

$$p_3(x)=1+\frac{1}{2}x-\frac{1}{8}x^2+\frac{1}{16}x^3 \quad .$$

Therefore,

$$\sqrt{1.5}=\sqrt{1+0.5}\approx 1+\frac{1}{2}(0.5)-\frac{1}{8}(0.5)^2+\frac{1}{16}(0.5)^3\approx 1.2266.$$

From Taylor's theorem, the error satisfies

$$R_3(0.5)=\frac{f^{(4)}(c)}{4!}(0.5)^4$$

for some c between 0 and 0.5. Since $f^{(4)}(x)=-\dfrac{15}{2^4(1+x)^{7/2}}$, and the maximum value of $\left|f^{(4)}(x)\right|$ on the interval $(0,0.5)$ occurs at $x=0$, we have

$$|R_3(0.5)|\le\frac{15}{4!2^4}(0.5)^4\approx 0.00244.$$

The function and the Maclaurin polynomial p_3 are graphed in Figure 6.4.1.

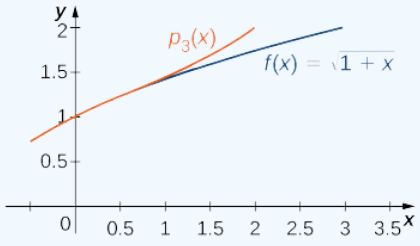


Figure 6.4.1: The third-order Maclaurin polynomial $p_3(x)$ provides a good approximation for $f(x)=\sqrt{1+x}$ for x near zero.

Exercise 6.4.1

Find the binomial series for $f(x)=\dfrac{1}{(1+x)^2}$.

Hint

Use the definition of binomial series for $r=-2$.

Answer

$$\sum_{n=0}^{\infty}(-1)^n(n+1)x^n$$

Common Functions Expressed as Taylor Series

At this point, we have derived Maclaurin series for exponential, trigonometric, and logarithmic functions, as well as functions of the form $f(x)=(1+x)^r$. In Table 6.4.1, we summarize the results of these series. We remark that the convergence of the Maclaurin series for $f(x)=\ln(1+x)$ at the endpoint $x=1$ and the Maclaurin series for $f(x)=\tan^{-1}x$ at the endpoints $x=1$ and $x=-1$ relies on a more advanced theorem than we present here. (Refer to Abel's theorem for a discussion of this more technical point.)

Table 6.4.1: Maclaurin Series for Common Functions

Function	Maclaurin Series	Interval of Convergence
$f(x)=\dfrac{1}{1-x}$	$\displaystyle\sum_{n=0}^{\infty}x^n$	$-1<x<1$
$f(x)=e^x$	$\displaystyle\sum_{n=0}^{\infty}\frac{x^n}{n!}$	$-\infty<x<\infty$
$f(x)=\sin x$	$\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}$	$-\infty<x<\infty$
$f(x)=\cos x$	$\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}$	$-\infty<x<\infty$
$f(x)=\ln(1+x)$	$\displaystyle\sum_{n=0}^{\infty}(-1)^{n+1}\frac{x^n}{n}$	$-1<x\le 1$
$f(x)=\tan^{-1}x$	$\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{2n+1}$	$-1\le x\le 1$
$f(x)=(1+x)^r$	$\displaystyle\sum_{n=0}^{\infty}\binom{r}{n}x^n$	$-1<x<1$ Convergence at the endpoints depends on the value of r

Earlier in the chapter, we showed how you could combine power series to create new power series. Here we use these properties, combined with the Maclaurin series in Table 6.4.1, to create Maclaurin series for other functions.

Example 6.4.2: Deriving Maclaurin Series from Known Series

Find the Maclaurin series of each of the following functions by using one of the series listed in Table 6.4.1.

a. $f(x)=\cos\sqrt{x}$
b. $f(x)=\sinh x$

Solution

a. Using the Maclaurin series for $\cos x$ we find that the Maclaurin series for $\cos\sqrt{x}$ is given by

$$\sum_{n=0}^{\infty}\frac{(-1)^n(\sqrt{x})^{2n}}{(2n)!}=\sum_{n=0}^{\infty}\frac{(-1)^nx^n}{(2n)!}=1-\frac{x}{2!}+\frac{x^2}{4!}-\frac{x^3}{6!}+\frac{x^4}{8!}-\cdots.$$

This series converges to $\cos\sqrt{x}$ for all x in the domain of $\cos\sqrt{x}$; that is, for all $x\ge 0$.

b. To find the Maclaurin series for $\sinh x$, we use the fact that

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

Using the Maclaurin series for e^x, we see that the nth term in the Maclaurin series for $\sinh x$ is given by

$$\frac{x^n}{n!} - \frac{(-x)^n}{n!}.$$

For n even, this term is zero. For n odd, this term is $\frac{2x^n}{n!}$. Therefore, the Maclaurin series for $\sinh x$ has only odd-order terms and is given by

$$\sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots.$$

? Exercise 6.4.2

Find the Maclaurin series for $\sin(x^2)$.

Hint

Use the Maclaurin series for $\sin x$.

Answer

$$\sum_{n=0}^{\infty} \frac{(-1)^n x^{4n+2}}{(2n+1)!}$$

We also showed previously in this chapter how power series can be differentiated term by term to create a new power series. In Example 6.4.3, we differentiate the binomial series for $\sqrt{1+x}$ term by term to find the binomial series for $\frac{1}{\sqrt{1+x}}$. Note that we could construct the binomial series for $\frac{1}{\sqrt{1+x}}$ directly from the definition, but differentiating the binomial series for $\sqrt{1+x}$ is an easier calculation.

✓ Example 6.4.3: Differentiating a Series to Find a New Series

Use the binomial series for $\sqrt{1+x}$ to find the binomial series for $\frac{1}{\sqrt{1+x}}$.

Solution

The two functions are related by

$$\frac{d}{dx}\sqrt{1+x} = \frac{1}{2\sqrt{1+x}},$$

so the binomial series for $\frac{1}{\sqrt{1+x}}$ is given by

$$\frac{1}{\sqrt{1+x}} = 2\frac{d}{dx}\sqrt{1+x} = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} x^n.$$

? Exercise 6.4.3

Find the binomial series for $f(x) = \frac{1}{(1+x)^{3/2}}$

Hint

Differentiate the series for $\frac{1}{\sqrt{1+x}}$

Answer

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1 \cdot 3 \cdot 5 \cdots (2n+1)}{2^n} x^n$$

In this example, we differentiated a known Taylor series to construct a Taylor series for another function. The ability to differentiate power series term by term makes them a powerful tool for solving differential equations. We now show how this is accomplished.

Solving Differential Equations with Power Series

Consider the differential equation

$$y'(x) = y.$$

Recall that this is a first-order separable equation and its solution is $y = Ce^x$. This equation is easily solved using techniques discussed earlier in the text. For most differential equations, however, we do not yet have analytical tools to solve them. Power series are an extremely useful tool for solving many types of differential equations. In this technique, we look for a solution of the form $y = \sum_{n=0}^{\infty} c_n x^n$ and determine what the coefficients would need to be. In the next example, we consider an initial-value problem involving $y' = y$ to illustrate the technique.

Example 6.4.4: Power Series Solution of a Differential Equation

Use power series to solve the initial-value problem $y' = y, \quad y(0) = 3.$

Solution

Suppose that there exists a power series solution

$$y(x) = \sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \cdots.$$

Differentiating this series term by term, we obtain

$$y' = c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 + \cdots.$$

If y satisfies the differential equation, then

$$c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots = c_1 + 2c_2 x + 3c_3 x^2 + 4c_3 x^3 + \cdots.$$

Using the uniqueness of power series representations, we know that these series can only be equal if their coefficients are equal. Therefore,

$$\begin{aligned} c_0 &= c_1, \\ c_1 &= 2c_2, \\ c_2 &= 3c_3, \\ c_3 &= 4c_4, \\ &\vdots \end{aligned}$$

Using the initial condition $y(0) = 3$ combined with the power series representation

$$y(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots \ ,$$

we find that $c_0 = 3$. We are now ready to solve for the rest of the coefficients. Using the fact that $c_0 = 3$, we have

$$
\begin{aligned}
c_1 &= c_0 = 3 = \frac{3}{1!},\\
c_2 &= \frac{c_1}{2} = \frac{3}{2} = \frac{3}{2!},\\
c_3 &= \frac{c_2}{3} = \frac{3}{3\cdot 2} = \frac{3}{3!},\\
c_4 &= \frac{c_3}{4} = \frac{3}{4\cdot 3\cdot 2} = \frac{3}{4!}.
\end{aligned}
$$

Therefore,

$$
y = 3\left[1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3\frac{1}{4!}x^4 + \cdots\right] = 3\sum_{n=0}^{\infty}\frac{x^n}{n!}.
$$

You might recognize

$$
\sum_{n=0}^{\infty}\frac{x^n}{n!}
$$

as the Taylor series for e^x. Therefore, the solution is $y = 3e^x$.

? Exercise 6.4.4

Use power series to solve $y' = 2y, \quad y(0) = 5.$

Hint

The equations for the first several coefficients c_n will satisfy $c_0 = 2c_1$, $c_1 = 2\cdot 2c_2$, $c_2 = 2\cdot 3c_3$, ... In general, for all $n \geq 0$, $c_n = 2(n+1)C_{n+1}$.

Answer

$y = 5e^{2x}$

We now consider an example involving a differential equation that we cannot solve using previously discussed methods. This differential equation

$$
y'' - xy = 0
$$

is known as **Airy's equation**. It has many applications in mathematical physics, such as modeling the diffraction of light. Here we show how to solve it using power series.

✓ Example 6.4.5: Power Series Solution of Airy's Equation

Use power series to solve $y'' - xy = 0$ with the initial conditions $y(0) = a$ and $y'(0) = b$.

Solution

We look for a solution of the form

$$
y = \sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \cdots
$$

Differentiating this function term by term, we obtain

$$
\begin{aligned}
y' &= c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 + \cdots,\\
y'' &= 2\cdot 1c_2 + 3\cdot 2c_3 x + 4\cdot 3c_4 x^2 + \cdots.
\end{aligned}
$$

If y satisfies the equation $y'' = xy$, then

$$2\cdot 1c_2+3\cdot 2c_3x+4\cdot 3c_4x^2+\cdots=x(c_0+c_1x+c_2x^2+c_3x^3+\cdots).$$

Using the Uniqueness of Power Series Theorem (from an earlier Section) on the uniqueness of power series representations, we know that coefficients of the same degree must be equal. Therefore,

$$\begin{aligned} 2\cdot 1c_2&=0,\\ 3\cdot 2c_3&=c_0,\\ 4\cdot 3c_4&=c_1,\\ 5\cdot 4c_5&=c_2,\\ &\vdots \end{aligned}$$

More generally, for $n\geq 3$, we have $n\cdot(n-1)c_n=c_{n-3}$. In fact, all coefficients can be written in terms of c_0 and c_1. To see this, first note that $c_2=0$. Then

$$\begin{aligned} c_3&=\frac{c_0}{3\cdot 2},\\ c_4&=\frac{c_1}{4\cdot 3}. \end{aligned}$$

For c_5, c_6, c_7, we see that

$$\begin{aligned} c_5&=\frac{c_2}{5\cdot 4}=0,\\ c_6&=\frac{c_3}{6\cdot 5}=\frac{c_0}{6\cdot 5\cdot 3\cdot 2},\\ c_7&=\frac{c_4}{7\cdot 6}=\frac{c_1}{7\cdot 6\cdot 4\cdot 3}. \end{aligned}$$

Therefore, the series solution of the differential equation is given by

$$y=c_0+c_1x+0\cdot x^2+\frac{c_0}{3\cdot 2}x^3+\frac{c_1}{4\cdot 3}x^4+0\cdot x^5+\frac{c_0}{6\cdot 5\cdot 3\cdot 2}x^6+\frac{c_1}{7\cdot 6\cdot 4\cdot 3}x^7+\cdots.$$

The initial condition $y(0)=a$ implies $c_0=a$. Differentiating this series term by term and using the fact that $y'(0)=b$, we conclude that $c_1=b$.

Therefore, the solution of this initial-value problem is

$$y=a\left(1+\frac{x^3}{3\cdot 2}+\frac{x^6}{6\cdot 5\cdot 3\cdot 2}+\cdots\right)+b\left(x+\frac{x^4}{4\cdot 3}+\frac{x^7}{7\cdot 6\cdot 4\cdot 3}+\cdots\right).$$

? Exercise 6.4.5

Use power series to solve $y''+x^2y=0$ with the initial condition $y(0)=a$ and $y'(0)=b$.

Hint

The coefficients satisfy $c_0=a$, $c_1=b$, $c_2=0$, $c_3=0$, and for $n\geq 4$, $n(n-1)c_n=-c_{n-4}$.

Answer

$$y=a\left(1-\frac{x^4}{3\cdot 4}+\frac{x^8}{3\cdot 4\cdot 7\cdot 8}-\cdots\right)+b\left(x-\frac{x^5}{4\cdot 5}+\frac{x^9}{4\cdot 5\cdot 8\cdot 9}-\cdots\right)$$

Evaluating Non-elementary Integrals

Solving differential equations is one common application of power series. We now turn to a second application. We show how power series can be used to evaluate integrals involving functions whose antiderivatives cannot be expressed using elementary functions.

One integral that arises often in applications in probability theory is $\int e^{-x^2}\,dx$. Unfortunately, the antiderivative of the integrand e^{-x^2} is not an elementary function. By elementary function, we mean a function that can be written using a finite number of algebraic combinations or compositions of exponential, logarithmic, trigonometric, or power functions. We remark that the term "elementary function" is not synonymous with noncomplicated function. For example, the function $f(x)=\sqrt{x^2-3x}+e^{x^3}-\sin(5x+4)$ is an elementary function, although not a particularly simple-looking function. Any integral of the form $\int f(x)\,dx$ where the antiderivative of f cannot be written as an elementary function is considered a **non-elementary integral.**

Non-elementary integrals cannot be evaluated using the basic integration techniques discussed earlier. One way to evaluate such integrals is by expressing the integrand as a power series and integrating term by term. We demonstrate this technique by considering $\int e^{-x^2}\,dx$.

Example 6.4.6: Using Taylor Series to Evaluate a Definite Integral

a. Express $\int e^{-x^2}\,dx$ as an infinite series.

b. Evaluate $\int_0^1 e^{-x^2}\,dx$ to within an error of 0.01.

Solution

a. The Maclaurin series for e^{-x^2} is given by

$$\begin{aligned} e^{-x^2} &= \sum_{n=0}^{\infty}\frac{(-x^2)^n}{n!} \\ &= 1-x^2+\frac{x^4}{2!}-\frac{x^6}{3!}+\cdots+(-1)^n\frac{x^{2n}}{n!}+\cdots \\ &= \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{n!}. \end{aligned}$$

Therefore,

$$\begin{aligned} \int e^{-x^2}\,dx &= \int\left(1-x^2+\frac{x^4}{2!}-\frac{x^6}{3!}+\cdots+(-1)^n\frac{x^{2n}}{n!}+\cdots\right)dx \\ &= C+x-\frac{x^3}{3}+\frac{x^5}{5.2!}-\frac{x^7}{7.3!}+\cdots+(-1)^n\frac{x^{2n+1}}{(2n+1)n!}+\cdots. \end{aligned}$$

b. Using the result from part a. we have

$$\int_0^1 e^{-x^2}\,dx = 1-\frac{1}{3}+\frac{1}{10}-\frac{1}{42}+\frac{1}{216}-\cdots.$$

The sum of the first four terms is approximately 0.74. By the alternating series test, this estimate is accurate to within an error of less than $\frac{1}{216}\approx 0.0046296<0.01$.

Exercise 6.4.6

Express $\int \cos\sqrt{x}\,dx$ as an infinite series. Evaluate $\int_0^1 \cos\sqrt{x}\,dx$ to within an error of 0.01.

Hint

Use the series found in Example $6.4.2$

Answer

$C + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n(2n-2)!}$ The definite integral is approximately 0.764 to within an error of 0.01.

As mentioned above, the integral $\int e^{-x^2}\,dx$ arises often in probability theory. Specifically, it is used when studying data sets that are normally distributed, meaning the data values lie under a bell-shaped curve. For example, if a set of data values is normally distributed with mean μ and standard deviation σ, then the probability that a randomly chosen value lies between $x = a$ and $x = b$ is given by

$$\frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-(x-\mu)^2/(2\sigma^2)}\,dx. \tag{6.4.5}$$

(See Figure 6.4.2.)

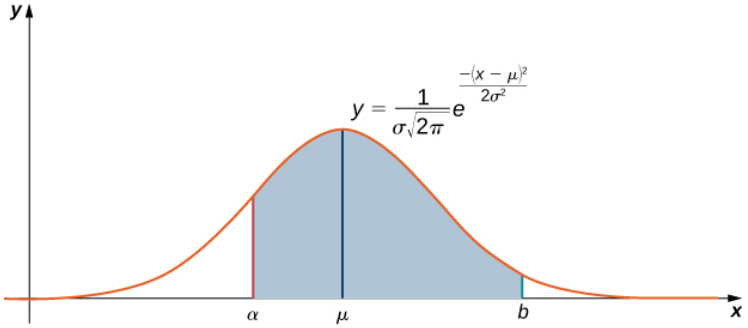


Figure 6.4.2: If data values are normally distributed with mean μ and standard deviation σ, the probability that a randomly selected data value is between a and b is the area under the curve $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ between $x = a$ and $x = b$.

To simplify this integral, we typically let $z = \frac{x-\mu}{\sigma}$. This quantity z is known as the z score of a data value. With this simplification, integral Equation 6.4.5 becomes

$$\frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} e^{-z^2/2}\,dz.$$

In Example 6.4.7, we show how we can use this integral in calculating probabilities.

Example 6.4.7: Using Maclaurin Series to Approximate a Probability

Suppose a set of standardized test scores are normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 50$. Use Equation 6.4.5 and the first six terms in the Maclaurin series for $e^{-x^2/2}$ to approximate the probability that a randomly selected test score is between $x = 100$ and $x = 200$. Use the alternating series test to determine how accurate your approximation is.

Solution

Since $\mu = 100, \sigma = 50$, and we are trying to determine the area under the curve from $a = 100$ to $b = 200$, integral Equation 6.4.5 becomes

$$\frac{1}{\sqrt{2\pi}} \int_0^2 e^{-z^2/2}\,dz. \tag{6.4.6}$$

The Maclaurin series for $e^{-x^2/2}$ is given by

$$e^{-x^2/2} = \sum_{n=0}^{\infty} \frac{\left(-\frac{x^2}{2}\right)^n}{n!}$$
$$= 1 - \frac{x^2}{2^1 \cdot 1!} + \frac{x^4}{2^2 \cdot 2!} - \frac{x^6}{2^3 \cdot 3!} + \cdots + (-1)^n \frac{x^{2n}}{2^n \cdot n}! + \cdots$$
$$= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{2^n \cdot n!}.$$

Therefore,

$$\frac{1}{\sqrt{2\pi}} \int e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \int \left(1 - \frac{z^2}{2^1 \cdot 1!} + \frac{z^4}{2^2 \cdot 2!} - \frac{z^6}{2^3 \cdot 3!} + \cdots + (-1)^n \frac{z^{2n}}{2^n \cdot n!} + \cdots\right) dz$$
$$= \frac{1}{\sqrt{2\pi}} \left(C + z - \frac{z^3}{3 \cdot 2^1 \cdot 1!} + \frac{z^5}{5 \cdot 2^2 \cdot 2!} - \frac{z^7}{7 \cdot 2^3 \cdot 3!} + \cdots + (-1)^n \frac{z^{2n+1}}{(2n+1)2^n \cdot n!} + \cdots\right)$$

We now use this result to evaluate the definite integral from Equation 6.4.6:

$$\frac{1}{\sqrt{2\pi}} \int_0^2 e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \left(2 - \frac{8}{6} + \frac{32}{40} - \frac{128}{336} + \frac{512}{3456} - \frac{2^{11}}{11 \cdot 2^5 \cdot 5!} + \cdots\right)$$

Using the first five terms, we estimate that the probability is approximately **0.4729.** By the alternating series test, we see that this estimate is accurate to within

$$\frac{1}{\sqrt{2\pi}} \frac{2^{13}}{13 \cdot 2^6 \cdot 6!} \approx 0.00546.$$

Analysis

If you are familiar with probability theory, you may know that the probability that a data value is within two standard deviations of the mean is approximately 95%. Here we calculated the probability that a data value is between the mean and two standard deviations above the mean, so the estimate should be around 47.5%. The estimate, combined with the bound on the accuracy, falls within this range.

Alternative method. If term-by-term integration is done instead, the result of evaluating the definite integral is

$$\frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \int_0^2 z^{2n}\, dz = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)2^n n!} \left[2^{2n+1} - 0^{2n+1}\right] = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n 2^{n+1}}{n!(2n+1)}$$
$$= \frac{1}{\sqrt{2\pi}} \left(\frac{2}{1 \cdot 0!} - \frac{4}{3 \cdot 1!} + \frac{8}{5 \cdot 2!} - \frac{16}{7 \cdot 3!} + \frac{32}{9 \cdot 4!} - \frac{64}{11 \cdot 5!} + \cdots\right)$$
$$= \frac{1}{\sqrt{2\pi}} \left(2 - \frac{4}{3} + \frac{4}{5} - \frac{8}{21} + \frac{4}{27} - \frac{8}{165} + \cdots\right)$$

? Exercise 6.4.7

Given a set of normally distributed standardized test scores with mean $\mu = 100$ and standard deviation $\sigma = 50$, use the first five terms of the Maclaurin series for $e^{-x^2/2}$ to estimate the probability that a randomly selected test score is between 100 and 150. Use the alternating series test to determine the accuracy of this estimate.

Hint

Evaluate $\int_0^1 e^{-z^2/2}\, dz$ using the first five terms of the Maclaurin series for $e^{-z^2/2}$.

Answer

The estimate is approximately 0.3414.This estimate is accurate to within 0.0000094.

Another application in which a non-elementary integral arises involves the period of a pendulum. The integral is

$$\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}$$

.

An integral of this form is known as an **elliptic integral** of the first kind. Elliptic integrals originally arose when trying to calculate the arc length of an ellipse. We now show how to use power series to approximate this integral.

Example 6.4.8: Period of a Pendulum

The period of a pendulum is the time it takes for a pendulum to make one complete back-and-forth swing. For a pendulum with length L that makes a maximum angle θ_{max} with the vertical, its period T is given by

$$T = 4\sqrt{\frac{L}{g}}\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}$$

where g is the acceleration due to gravity and $k = \sin\left(\frac{\theta_{max}}{2}\right)$ (see Figure 6.4.3). (We note that this formula for the period arises from a non-linearized model of a pendulum. In some cases, for simplification, a linearized model is used and $\sin\theta$ is approximated by θ.)

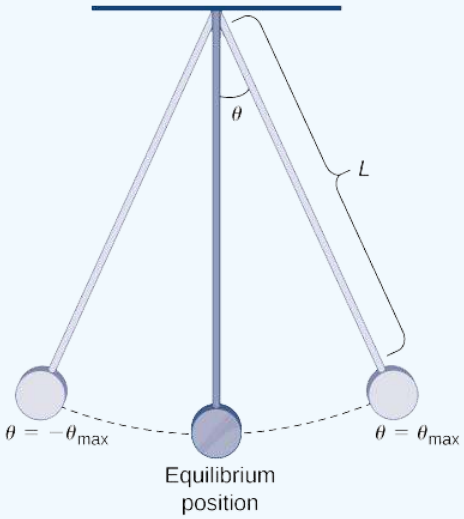


Figure 6.4.3: This pendulum has length L and makes a maximum angle θ_{max} with the vertical.

Use the binomial series

$$\frac{1}{\sqrt{1+x}} = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n}x^n$$

to estimate the period of this pendulum. Specifically, approximate the period of the pendulum if

a. you use only the first term in the binomial series, and
b. you use the first two terms in the binomial series.

Solution

We use the binomial series, replacing x with $-k^2\sin^2\theta$. Then we can write the period as

$$T = 4\sqrt{\frac{L}{g}}\int_0^{\pi/2}\left(1+\frac{1}{2}k^2\sin^2\theta+\frac{1\cdot 3}{2!2^2}k^4\sin^4\theta+\cdots\right)d\theta.$$

a. Using just the first term in the integrand, the first-order estimate is

$$T \approx 4\sqrt{\frac{L}{g}}\int_0^{\pi/2} d\theta = 2\pi\sqrt{\frac{L}{g}}.$$

If θ_{max} is small, then $k = \sin\left(\frac{\theta_{max}}{2}\right)$ is small. We claim that when k is small, this is a good estimate. To justify this claim, consider

$$\int_0^{\pi/2}\left(1+\frac{1}{2}k^2\sin^2\theta+\frac{1\cdot 3}{2!2^2}k^4\sin^4\theta+\cdots\right)d\theta.$$

Since $|\sin x| \leq 1$, this integral is bounded by

$$\int_0^{\pi/2}\left(\frac{1}{2}k^2+\frac{1.3}{2!2^2}k^4+\cdots\right)d\theta \;<\; \frac{\pi}{2}\left(\frac{1}{2}k^2+\frac{1\cdot 3}{2!2^2}k^4+\cdots\right).$$

Furthermore, it can be shown that each coefficient on the right-hand side is less than 1 and, therefore, that this expression is bounded by

$$\frac{\pi k^2}{2}(1+k^2+k^4+\cdots) = \frac{\pi k^2}{2}\cdot\frac{1}{1-k^2} \quad ,$$

which is small for k small.

b. For larger values of θ_{max}, we can approximate T by using more terms in the integrand. By using the first two terms in the integral, we arrive at the estimate

$$T \approx 4\sqrt{\frac{L}{g}}\int_0^{\pi/2}\left(1+\frac{1}{2}k^2\sin^2\theta\right)d\theta = 2\pi\sqrt{\frac{L}{g}}\left(1+\frac{k^2}{4}\right).$$

The applications of Taylor series in this section are intended to highlight their importance. In general, Taylor series are useful because they allow us to represent known functions using polynomials, thus providing us a tool for approximating function values and estimating complicated integrals. In addition, they allow us to define new functions as power series, thus providing us with a powerful tool for solving differential equations.

Key Concepts

- The binomial series is the Maclaurin series for $f(x) = (1+x)^r$. It converges for $|x| < 1$.
- Taylor series for functions can often be derived by algebraic operations with a known Taylor series or by differentiating or integrating a known Taylor series.
- Power series can be used to solve differential equations.
- Taylor series can be used to help approximate integrals that cannot be evaluated by other means.

Glossary

binomial series

the Maclaurin series for $f(x) = (1+x)^r$; it is given by

$$(1+x)^r = \sum_{n=0}^{\infty}\binom{r}{n}x^n = 1+rx+\frac{r(r-1)}{2!}x^2+\cdots+\frac{r(r-1)\cdots(r-n+1)}{n!}x^n+\cdots \qquad \text{for } |x|<1$$

non-elementary integral

an integral for which the antiderivative of the integrand cannot be expressed as an elementary function

This page titled 6.4: Working with Taylor Series is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.4: Working with Taylor Series** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

CHAPTER OVERVIEW

7: Parametric Equations and Polar Coordinates

Parametric equations define a group of quantities as functions of one or more independent variables called parameters. Parametric equations are commonly used to express the coordinates of the points that make up a geometric object such as a curve or surface, in which case the equations are collectively called a parametric representation or parameterization. The polar coordinate system is a two-dimensional coordinate system in which each point on a plane is determined by a distance from a reference point and an angle from a reference direction. The reference point (analogous to the origin of a Cartesian system) is called the pole, and the ray from the pole in the reference direction is the polar axis. The distance from the pole is called the radial coordinate or radius, and the angle is called the angular coordinate, polar angle, or azimuth

7.E: Parametric Equations and Polar Coordinates (Exercises)
7.0: Prelude to Parametric Equations and Polar Coordinates
7.1: Parametric Equations
7.2: Calculus of Parametric Curves
7.3: Polar Coordinates
7.4: Area and Arc Length in Polar Coordinates
7.5: Conic Sections

Contributors

Gilbert Strang (MIT) and Edwin “Jed” Herman (Harvey Mudd) with many contributing authors. This content by OpenStax is licensed with a CC-BY-SA-NC 4.0 license. Download for free at http://cnx.org.

This page titled 7: Parametric Equations and Polar Coordinates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

7.E: Parametric Equations and Polar Coordinates (Exercises)

This page titled 7.E: Parametric Equations and Polar Coordinates (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

7.0: Prelude to Parametric Equations and Polar Coordinates

The **chambered nautilus** is a fascinating creature. This animal feeds on hermit crabs, fish, and other crustaceans. It has a hard outer shell with many chambers connected in a spiral fashion, and it can retract into its shell to avoid predators. When part of the shell is cut away, a perfect spiral is revealed, with chambers inside that are somewhat similar to growth rings in a tree.

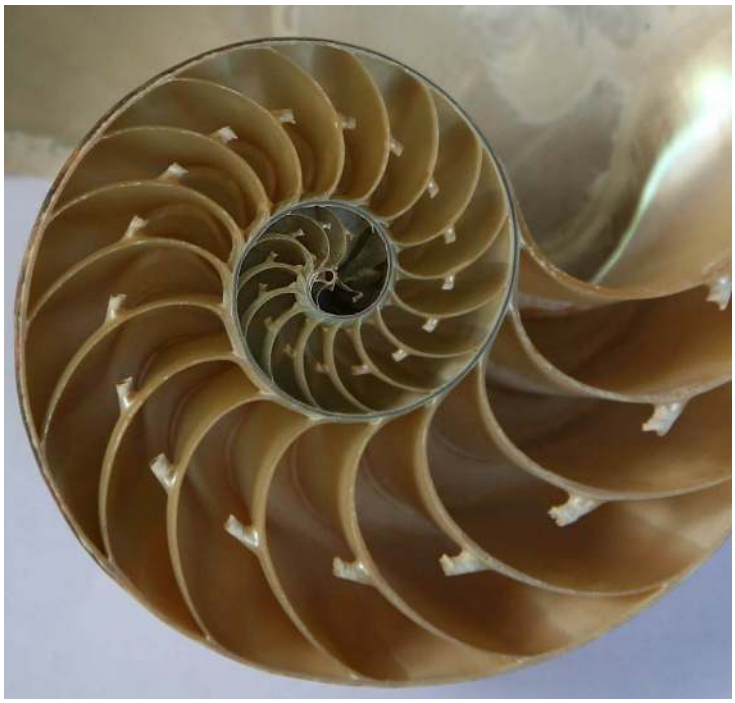


Figure 7.0.1: The chambered nautilus is a marine animal that lives in the tropical Pacific Ocean. Scientists think they have existed mostly unchanged for about 500 million years.(credit: modification of work by Jitze Couperus, Flickr)

The mathematical function that describes a spiral can be expressed using rectangular (or Cartesian) coordinates. However, if we change our coordinate system to something that works a bit better with circular patterns, the function becomes much simpler to describe. The polar coordinate system is well suited for describing curves of this type. How can we use this coordinate system to describe spirals and other radial figures?

In this chapter we also study parametric equations, which give us a convenient way to describe curves, or to study the position of a particle or object in two dimensions as a function of time. We will use parametric equations and polar coordinates for describing many topics later in this text.

This page titled 7.0: Prelude to Parametric Equations and Polar Coordinates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **11.0: Prelude to Parametric Equations and Polar Coordinates** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

7.1: Parametric Equations

Learning Objectives

- Plot a curve described by parametric equations.
- Convert the parametric equations of a curve into the form $y = f(x)$.
- Recognize the parametric equations of basic curves, such as a line and a circle.
- Recognize the parametric equations of a cycloid.

In this section we examine parametric equations and their graphs. In the two-dimensional coordinate system, parametric equations are useful for describing curves that are not necessarily functions. The parameter is an independent variable that both x and y depend on, and as the parameter increases, the values of x and y trace out a path along a plane curve. For example, if the parameter is t (a common choice), then t might represent time. Then x and y are defined as functions of time, and $(x(t), y(t))$ can describe the position in the plane of a given object as it moves along a curved path.

Parametric Equations and Their Graphs

Consider the orbit of Earth around the Sun. Our year lasts approximately 365.25 days, but for this discussion we will use 365 days. On January 1 of each year, the physical location of Earth with respect to the Sun is nearly the same, except for leap years, when the lag introduced by the extra $\frac{1}{4}$ day of orbiting time is built into the calendar. We call January 1 "day 1" of the year. Then, for example, day 31 is January 31, day 59 is February 28, and so on.

The number of the day in a year can be considered a variable that determines Earth's position in its orbit. As Earth revolves around the Sun, its physical location changes relative to the Sun. After one full year, we are back where we started, and a new year begins. According to Kepler's laws of planetary motion, the shape of the orbit is elliptical, with the Sun at one focus of the ellipse. We study this idea in more detail in Conic Sections.

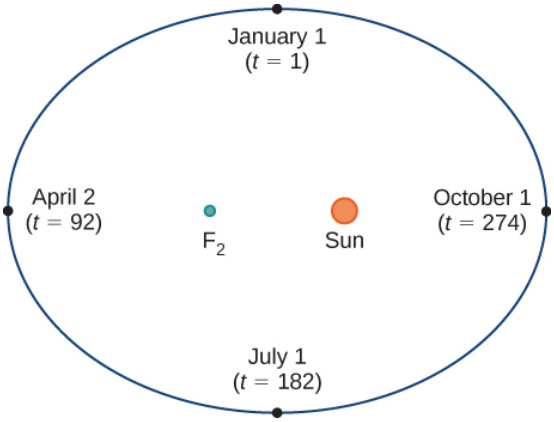


Figure 7.1.1: Earth's orbit around the Sun in one year.

Figure 7.1.1 depicts **Earth's orbit** around the Sun during one year. The point labeled F_2 is one of the foci of the ellipse; the other focus is occupied by the Sun. If we superimpose coordinate axes over this graph, then we can assign ordered pairs to each point on the ellipse (Figure 7.1.2). Then each x value on the graph is a value of position as a function of time, and each y value is also a value of position as a function of time. Therefore, each point on the graph corresponds to a value of Earth's position as a function of time.

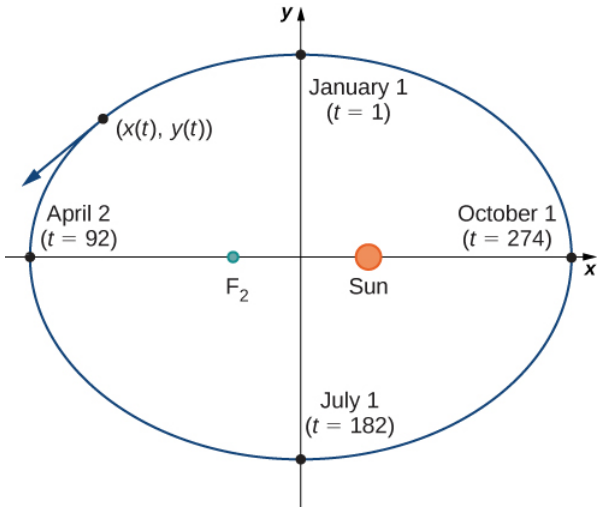


Figure 7.1.2: Coordinate axes superimposed on the orbit of Earth.

We can determine the functions for $x(t)$ and $y(t)$, thereby parameterizing the orbit of Earth around the Sun. The variable t is called an independent parameter and, in this context, represents time relative to the beginning of each year.

A curve in the (x, y) plane can be represented parametrically. The equations that are used to define the curve are called *parametric equations.*

Definition: Parametric Equations

If x and y are continuous functions of t on an interval I, then the equations

$$x = x(t)$$

and

$$y = y(t)$$

are called parametric equations and t is called the parameter. The set of points (x, y) obtained as t varies over the interval I is called the graph of the parametric equations. The graph of parametric equations is called a *parametric curve* or plane curve, and is denoted by C.

Notice in this definition that x and y are used in two ways. The first is as functions of the independent variable t. As t varies over the interval I, the functions $x(t)$ and $y(t)$ generate a set of ordered pairs (x, y). This set of ordered pairs generates the graph of the parametric equations. In this second usage, to designate the ordered pairs, x and y are variables. It is important to distinguish the variables x and y from the functions $x(t)$ and $y(t)$.

Example 7.1.1: Graphing a Parametrically Defined Curve

Sketch the curves described by the following parametric equations:

a. $x(t) = t - 1, \quad y(t) = 2t + 4, \quad \text{for } -3 \le t \le 2$
b. $x(t) = t^2 - 3, \quad y(t) = 2t + 1, \quad \text{for } -2 \le t \le 3$
c. $x(t) = 4\cos t, \quad y(t) = 4\sin t, \quad \text{for } 0 \le t \le 2\pi$

Solution a

To create a graph of this curve, first set up a table of values. Since the independent variable in both $x(t)$ and $y(t)$ is t, let t appear in the first column. Then $x(t)$ and $y(t)$ will appear in the second and third columns of the table.

t	$x(t)$	$y(t)$
−3	−4	−2
−2	−3	0

t	$x(t)$	$y(t)$
−1	−2	2
0	−1	4
1	0	6
2	1	8

The second and third columns in this table provide a set of points to be plotted. The graph of these points appears in Figure 7.1.3. The arrows on the graph indicate the **orientation** of the graph, that is, the direction that a point moves on the graph as t varies from −3 to 2.

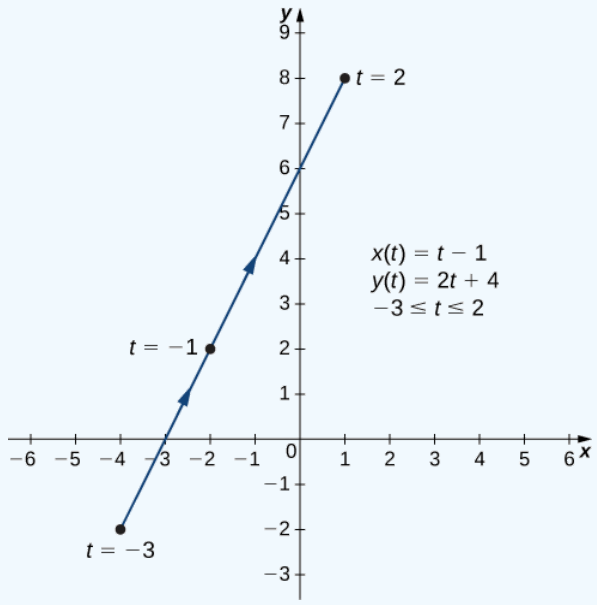


Figure 7.1.3: Graph of the plane curve described by the parametric equations in part a.

Solution b

To create a graph of this curve, again set up a table of values.

t	$x(t)$	$y(t)$
−2	1	−3
−1	−2	−1
0	−3	1
1	−2	3
2	1	5
3	6	7

The second and third columns in this table give a set of points to be plotted (Figure 7.1.4). The first point on the graph (corresponding to $t = -2$) has coordinates $(1, -3)$, and the last point (corresponding to $t = 3$) has coordinates $(6, 7)$. As t progresses from -2 to 3, the point on the curve travels along a parabola. The direction the point moves is again called the orientation and is indicated on the graph.

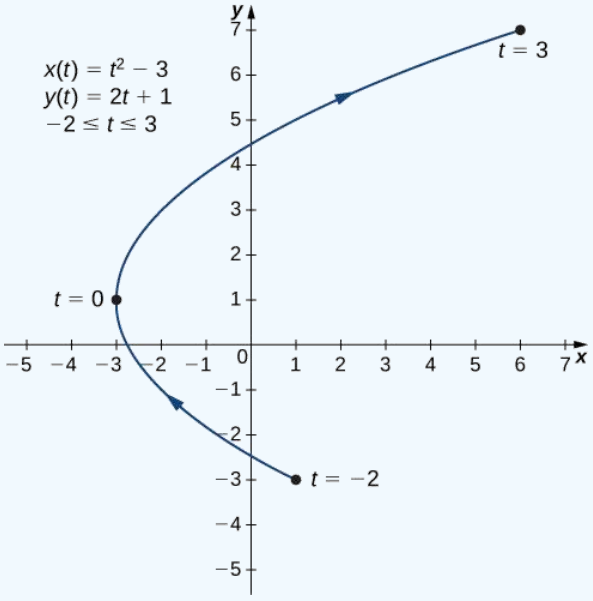


Figure 7.1.4: Graph of the plane curve described by the parametric equations in part b.

Solution c

In this case, use multiples of $\pi/6$ for t and create another table of values:

t	$x(t)$	$y(t)$	t	$x(t)$	$y(t)$
0	4	0	$\frac{7\pi}{6}$	$-2\sqrt{3} \approx -3.5$	-2
$\frac{\pi}{6}$	$2\sqrt{3} \approx 3.5$	2	$\frac{4\pi}{3}$	-2	$-2\sqrt{3} \approx -3.5$
$\frac{\pi}{3}$	2	$2\sqrt{3} \approx 3.5$	$\frac{3\pi}{2}$	0	-4
$\frac{\pi}{2}$	0	4	$\frac{5\pi}{3}$	2	$-2\sqrt{3} \approx -3.5$
$\frac{2\pi}{3}$	-2	$2\sqrt{3} \approx 3.5$	$\frac{11\pi}{6}$	$2\sqrt{3} \approx 3.5$	-2
$\frac{5\pi}{6}$	$-2\sqrt{3} \approx -3.5$	2	2π	4	0
π	-4	0			

The graph of this plane curve appears in the following graph.

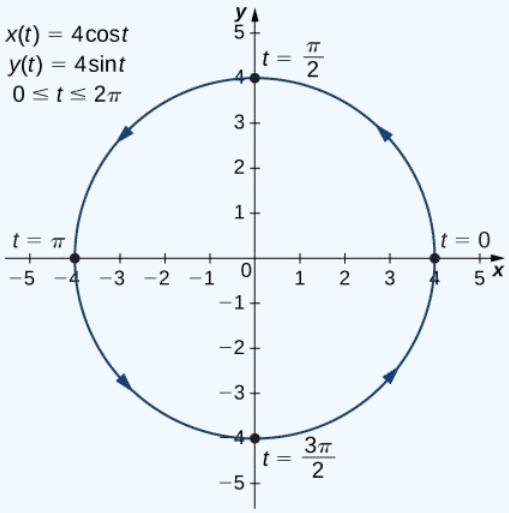


Figure 7.1.5: Graph of the plane curve described by the parametric equations in part c.

This is the graph of a circle with radius 4 centered at the origin, with a counterclockwise orientation. The starting point and ending points of the curve both have coordinates $(4, 0)$.

? Exercise 7.1.1

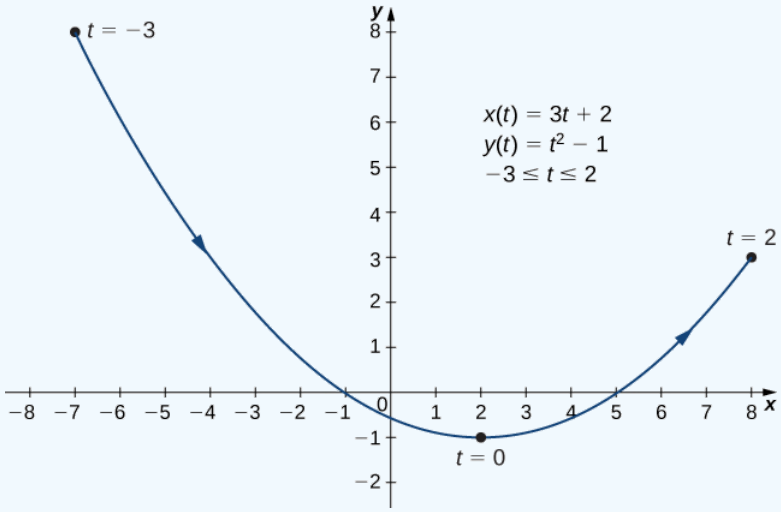
Sketch the curve described by the parametric equations

$$x(t) = 3t + 2, \quad y(t) = t^2 - 1, \quad \text{for } -3 \le t \le 2.$$

Hint

Make a table of values for $x(t)$ and $y(t)$ using t values from -3 to 2.

Answer



Eliminating the Parameter

To better understand the graph of a curve represented parametrically, it is useful to rewrite the two equations as a single equation relating the variables x and y. Then we can apply any previous knowledge of equations of curves in the plane to identify the curve. For example, the equations describing the plane curve in Example $7.1.1b$ are

$$x(t) = t^2 - 3 \tag{7.1.1}$$

$$y(t) = 2t + 1 \tag{7.1.2}$$

over the region $-2 \le t \le 3$.

Solving Equation 7.1.2 for t gives

$$t = \frac{y-1}{2}.$$

This can be substituted into Equation 7.1.1:

$$x = \left(\frac{y-1}{2}\right)^2 - 3 \tag{7.1.3}$$

$$= \frac{y^2 - 2y + 1}{4} - 3 \tag{7.1.4}$$

$$= \frac{y^2 - 2y - 11}{4}. \tag{7.1.5}$$

Equation 7.1.5 describes x as a function of y. These steps give an example of eliminating the parameter. The graph of this function is a parabola opening to the right (Figure 7.1.4). Recall that the plane curve started at $(1,-3)$ and ended at $(6,7)$. These terminations were due to the restriction on the parameter t.

Example 7.1.2: Eliminating the Parameter

Eliminate the parameter for each of the plane curves described by the following parametric equations and describe the resulting graph.

a. $x(t)=\sqrt{2t+4}, \quad y(t)=2t+1, \quad \text{for } -2\le t\le 6$
b. $x(t)=4\cos t, \quad y(t)=3\sin t, \quad \text{for } 0\le t\le 2\pi$

Solution a

To eliminate the parameter, we can solve either of the equations for t. For example, solving the first equation for t gives

$$\begin{aligned} x &= \sqrt{2t+4} \\ x^2 &= 2t+4 \\ x^2-4 &= 2t \\ t &= \frac{x^2-4}{2}. \end{aligned}$$

Note that when we square both sides it is important to observe that $x\ge 0$. Substituting $t=\dfrac{x^2-4}{2}$ into $y(t)$ yields

$$\begin{aligned} y(t) &= 2t+1 \\ y &= 2\left(\frac{x^2-4}{2}\right)+1 \\ y &= x^2-4+1 \\ y &= x^2-3. \end{aligned}$$

This is the equation of a parabola opening upward. There is, however, a domain restriction because of the limits on the parameter t. When $t=-2$, $x=\sqrt{2(-2)+4}=0$, and when $t=6$, $x=\sqrt{2(6)+4}=4$. The graph of this plane curve follows.

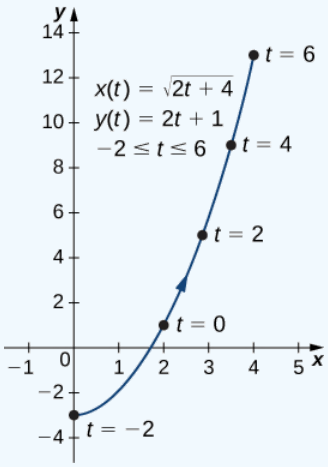


Figure 7.1.6: Graph of the plane curve described by the parametric equations in part a.

Solution b

Sometimes it is necessary to be a bit creative in eliminating the parameter. The parametric equations for this example are

$$x(t)=4\cos t$$

and

$$y(t) = 3\sin t$$

Solving either equation for t directly is not advisable because sine and cosine are not one-to-one functions. However, dividing the first equation by 4 and the second equation by 3 (and suppressing the t) gives us

$$\cos t = \frac{x}{4}$$

and

$$\sin t = \frac{y}{3}.$$

Now use the Pythagorean identity $\cos^2 t + \sin^2 t = 1$ and replace the expressions for $\sin t$ and $\cos t$ with the equivalent expressions in terms of x and y. This gives

$$\left(\frac{x}{4}\right)^2 + \left(\frac{y}{3}\right)^2 = 1$$

$$\frac{x^2}{16} + \frac{y^2}{9} = 1.$$

This is the equation of a horizontal ellipse centered at the origin, with semi-major axis 4 and semi-minor axis 3 as shown in the following graph.

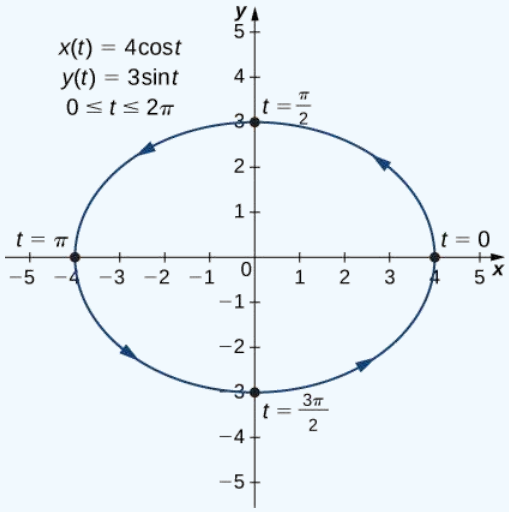


Figure 7.1.7: Graph of the plane curve described by the parametric equations in part b.

As t progresses from 0 to 2π, a point on the curve traverses the ellipse once, in a counterclockwise direction. Recall from the section opener that the orbit of Earth around the Sun is also elliptical. This is a perfect example of using parameterized curves to model a real-world phenomenon.

Exercise 7.1.2

Eliminate the parameter for the plane curve defined by the following parametric equations and describe the resulting graph.

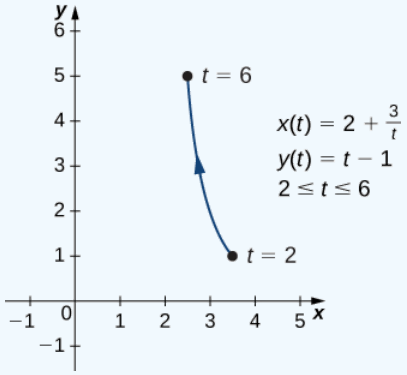
$$x(t) = 2 + \frac{3}{t}, \quad y(t) = t - 1, \quad \text{for } 2 \le t \le 6$$

Hint

Solve one of the equations for t and substitute into the other equation.

Answer

$x = 2 + \frac{3}{y+1}$, or $y = -1 + \frac{3}{x-2}$. This equation describes a portion of a rectangular hyperbola centered at $(2, -1)$.



So far we have seen the method of eliminating the parameter, assuming we know a set of parametric equations that describe a plane curve. What if we would like to start with the equation of a curve and determine a pair of parametric equations for that curve? This is certainly possible, and in fact it is possible to do so in many different ways for a given curve. The process is known as **parameterization of a curve.**

Example 7.1.3: Parameterizing a Curve

Find two different pairs of parametric equations to represent the graph of $y = 2x^2 - 3$.

Solution

First, it is always possible to parameterize a curve by defining $x(t) = t$, then replacing x with t in the equation for $y(t)$. This gives the parameterization

$$x(t) = t, \quad y(t) = 2t^2 - 3.$$

Since there is no restriction on the domain in the original graph, there is no restriction on the values of t.

We have complete freedom in the choice for the second parameterization. For example, we can choose $x(t) = 3t - 2$. The only thing we need to check is that there are no restrictions imposed on x; that is, the range of $x(t)$ is all real numbers. This is the case for $x(t) = 3t - 2$. Now since $y = 2x^2 - 3$, we can substitute $x(t) = 3t - 2$ for x. This gives

$$y(t) = 2(3t-2)^2 - 2 = 2(9t^2 - 12t + 4) - 2 = 18t^2 - 24t + 8 - 2 = 18t^2 - 24t + 6.$$

Therefore, a second parameterization of the curve can be written as

$$x(t) = 3t - 2 \text{ and } y(t) = 18t^2 - 24t + 6.$$

Exercise 7.1.3

Find two different sets of parametric equations to represent the graph of $y = x^2 + 2x$.

Hint

Follow the steps in Example 7.1.3 Remember we have freedom in choosing the parameterization for $x(t)$.

Answer

One possibility is $x(t) = t, \quad y(t) = t^2 + 2t.$ Another possibility is $x(t) = 2t - 3, \quad y(t) = (2t-3)^2 + 2(2t-3) = 4t^2 - 8t + 3.$ There are, in fact, an infinite number of possibilities.

Cycloids and Other Parametric Curves

Imagine going on a bicycle ride through the country. The tires stay in contact with the road and rotate in a predictable pattern. Now suppose a very determined ant is tired after a long day and wants to get home. So he hangs onto the side of the tire and gets a free ride. The path that this ant travels down a straight road is called a *cycloid* (Figure 7.1.8). A cycloid generated by a circle (or bicycle wheel) of radius a is given by the parametric equations

$$x(t) = a(t - \sin t), \quad y(t) = a(1 - \cos t).$$

To see why this is true, consider the path that the center of the wheel takes. The center moves along the x-axis at a constant height equal to the radius of the wheel. If the radius is a, then the coordinates of the center can be given by the equations

$$x(t) = at, \quad y(t) = a$$

for any value of t. Next, consider the ant, which rotates around the center along a circular path. If the bicycle is moving from left to right then the wheels are rotating in a clockwise direction. A possible parameterization of the circular motion of the ant (relative to the center of the wheel) is given by

$$x(t) = -a \sin t$$
$$y(t) = -a \cos t.$$

(The negative sign is needed to reverse the orientation of the curve. If the negative sign were not there, we would have to imagine the wheel rotating counterclockwise.) Adding these equations together gives the equations for the cycloid.

$$x(t) = a(t - \sin t)$$
$$y(t) = a(1 - \cos t)$$

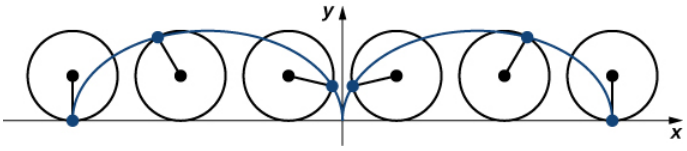


Figure 7.1.8: A wheel traveling along a road without slipping; the point on the edge of the wheel traces out a cycloid.

Now suppose that the bicycle wheel doesn't travel along a straight road but instead moves along the inside of a larger wheel, as in Figure 7.1.9. In this graph, the green circle is traveling around the blue circle in a counterclockwise direction. A point on the edge of the green circle traces out the red graph, which is called a *hypocycloid*.

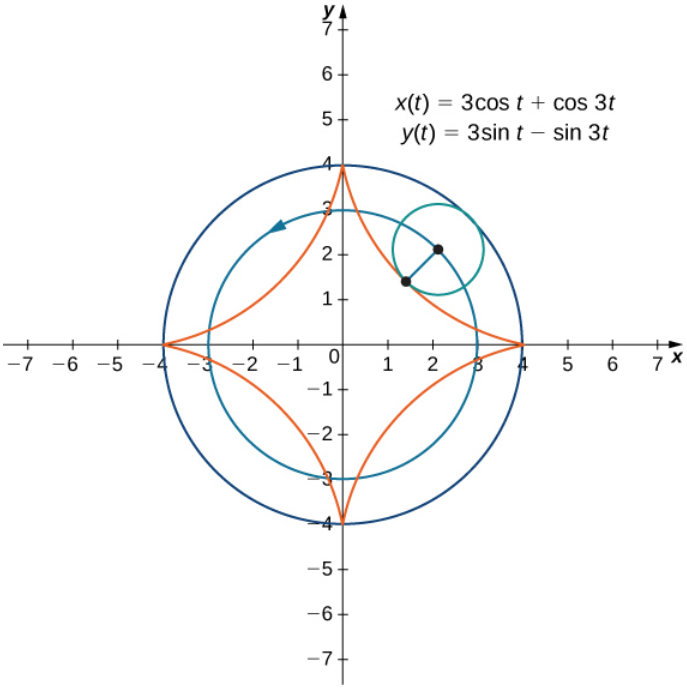


Figure 7.1.9: Graph of the hypocycloid described by the parametric equations shown.

The general parametric equations for a hypocycloid are

$$x(t) = (a - b)\cos t + b\cos\left(\frac{a - b}{b}\right)t$$

$$y(t) = (a-b)\sin t - b\sin(\frac{a-b}{b})t.$$

These equations are a bit more complicated, but the derivation is somewhat similar to the equations for the cycloid. In this case we assume the radius of the larger circle is a and the radius of the smaller circle is b. Then the center of the wheel travels along a circle of radius $a-b$. This fact explains the first term in each equation above. The period of the second trigonometric function in both $x(t)$ and $y(t)$ is equal to $\dfrac{2\pi b}{a-b}$.

The ratio $\dfrac{a}{b}$ is related to the number of **cusps** on the graph (cusps are the corners or pointed ends of the graph), as illustrated in Figure 7.1.10. This ratio can lead to some very interesting graphs, depending on whether or not the ratio is rational. Figure 7.1.9 corresponds to $a = 4$ and $b = 1$. The result is a hypocycloid with four cusps. Figure 7.1.10 shows some other possibilities. The last two hypocycloids have irrational values for $\dfrac{a}{b}$. In these cases the hypocycloids have an infinite number of cusps, so they never return to their starting point. These are examples of what are known as **space-filling curves.**

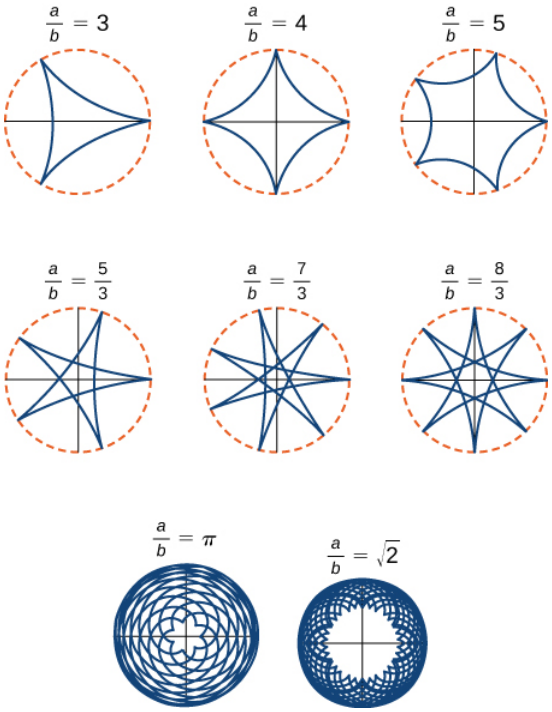


Figure 7.1.10: Graph of various hypocycloids corresponding to different values of a/b.

The Witch of Agnesi

Many plane curves in mathematics are named after the people who first investigated them, like the folium of Descartes or the spiral of Archimedes. However, perhaps the strangest name for a curve is the **witch of Agnesi.** Why a witch?

Maria Gaetana Agnesi (1718–1799) was one of the few recognized women mathematicians of eighteenth-century Italy. She wrote a popular book on analytic geometry, published in 1748, which included an interesting curve that had been studied by Fermat in 1630. The mathematician Guido Grandi showed in 1703 how to construct this curve, which he later called the "versoria," a Latin term for a rope used in sailing. Agnesi used the Italian term for this rope, "versiera," but in Latin, this same word means a "female goblin." When Agnesi's book was translated into English in 1801, the translator used the term "witch" for the curve, instead of rope. The name "witch of Agnesi" has stuck ever since.

The witch of Agnesi is a curve defined as follows: Start with a circle of radius a so that the points $(0, 0)$ and $(0, 2a)$ are points on the circle (Figure 7.1.11). Let O denote the origin. Choose any other point A on the circle, and draw the secant line OA. Let B denote the point at which the line OA intersects the horizontal line through $(0, 2a)$. The vertical line through B intersects the horizontal line through A at the point P. As the point A varies, the path that the point P travels is the witch of Agnesi curve for the given circle.

Witch of Agnesi curves have applications in physics, including modeling water waves and distributions of spectral lines. In probability theory, the curve describes the probability density function of the *Cauchy distribution*. In this project you will parameterize these curves.

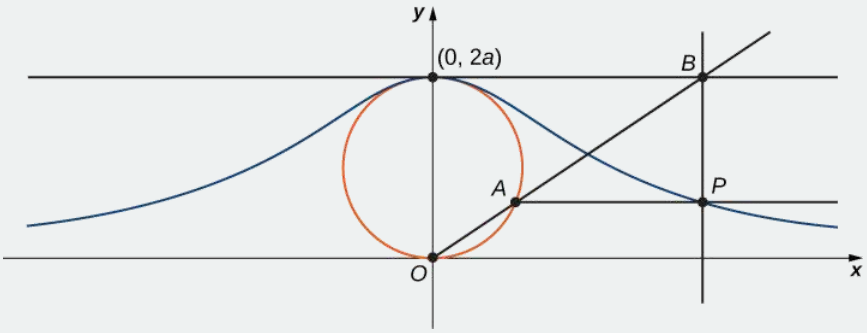


Figure 7.1.11: As the point A moves around the circle, the point P traces out the witch of Agnesi curve for the given circle.

1. On the figure, label the following points, lengths, and angle:
   a. C is the point on the x-axis with the same x-coordinate as A.
   b. x is the x-coordinate of P, and y is the y-coordinate of P.
   c. E is the point $(0, a)$.
   d. F is the point on the line segment OA such that the line segment EF is perpendicular to the line segment OA.
   e. b is the distance from O to F.
   f. c is the distance from F to A.
   g. d is the distance from O to B.
   h. θ is the measure of angle $\angle COA$.

The goal of this project is to parameterize the witch using θ as a parameter. To do this, write equations for x and y in terms of only θ.

2. Show that $d = \dfrac{2a}{\sin\theta}$.
3. Note that $x = d\cos\theta$. Show that $x = 2a\cot\theta$. When you do this, you will have parameterized the x-coordinate of the curve with respect to θ. If you can get a similar equation for y, you will have parameterized the curve.
4. In terms of θ, what is the angle $\angle EOA$?
5. Show that $b + c = 2a\cos\left(\frac{\pi}{2} - \theta\right)$.
6. Show that $y = 2a\cos\left(\frac{\pi}{2} - \theta\right)\sin\theta$.
7. Show that $y = 2a\sin^2\theta$. You have now parameterized the y-coordinate of the curve with respect to θ.
8. Conclude that a parameterization of the given witch curve is

$$x = 2a\cot\theta, \quad y = 2a\sin^2\theta, \quad \text{for } -\infty < \theta < \infty.$$

9. Use your parameterization to show that the given witch curve is the graph of the function $f(x) = \dfrac{8a^3}{x^2 + 4a^2}$.

Travels with My Ant: The Curtate and Prolate Cycloids

Earlier in this section, we looked at the parametric equations for a cycloid, which is the path a point on the edge of a wheel traces as the wheel rolls along a straight path. In this project we look at two different variations of the cycloid, called the curtate and prolate cycloids.

First, let's revisit the derivation of the parametric equations for a cycloid. Recall that we considered a tenacious ant trying to get home by hanging onto the edge of a bicycle tire. We have assumed the ant climbed onto the tire at the very edge, where the tire touches the ground. As the wheel rolls, the ant moves with the edge of the tire (Figure 7.1.12).

As we have discussed, we have a lot of flexibility when parameterizing a curve. In this case we let our parameter t represent the angle the tire has rotated through. Looking at Figure 7.1.12, we see that after the tire has rotated through an angle of t, the position of the center of the wheel, $C = (x_C, y_C)$, is given by

$$x_C = at \text{ and } y_C = a.$$

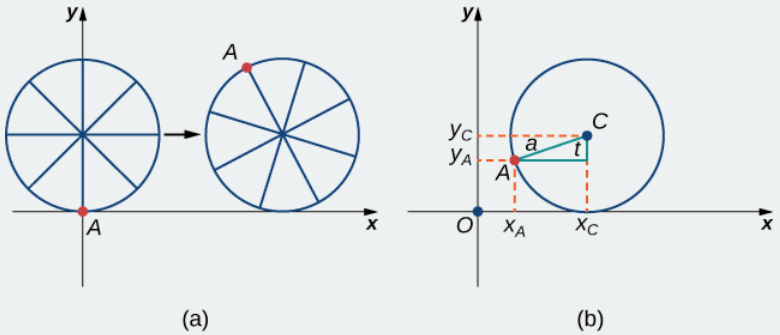
Furthermore, letting $A = (x_A, y_A)$ denote the position of the ant, we note that

$$x_C - x_A = a\sin t \text{ and } y_C - y_A = a\cos t$$

Then

$$x_A = x_C - a\sin t = at - a\sin t = a(t - \sin t)$$

$$y_A = y_C - a\cos t = a - a\cos t = a(1 - \cos t).$$

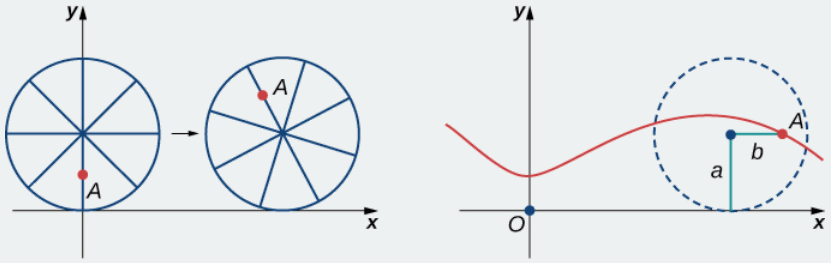


(a) (b)

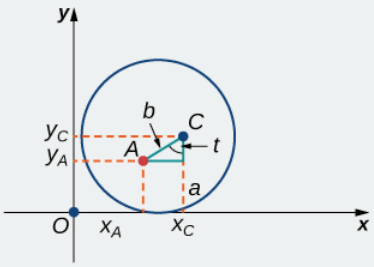
Figure 7.1.12: (a) The ant clings to the edge of the bicycle tire as the tire rolls along the ground. (b) Using geometry to determine the position of the ant after the tire has rotated through an angle of t.

Note that these are the same parametric representations we had before, but we have now assigned a physical meaning to the parametric variable t.

After a while the ant is getting dizzy from going round and round on the edge of the tire. So he climbs up one of the spokes toward the center of the wheel. By climbing toward the center of the wheel, the ant has changed his path of motion. The new path has less up-and-down motion and is called a **curtate cycloid** (Figure 7.1.13). As shown in the figure, we let b denote the distance along the spoke from the center of the wheel to the ant. As before, we let t represent the angle the tire has rotated through. Additionally, we let $C = (x_C, y_C)$ represent the position of the center of the wheel and $A = (x_A, y_A)$ represent the position of the ant.



(a) (b)



(c)

Figure 7.1.13: (a) The ant climbs up one of the spokes toward the center of the wheel. (b) The ant's path of motion after he climbs closer to the center of the wheel. This is called a curtate cycloid. (c) The new setup, now that the ant has moved closer to the center of the wheel.

1. What is the position of the center of the wheel after the tire has rotated through an angle of t?

2. Use geometry to find expressions for $x_C - x_A$ and for $y_C - y_A$.
3. On the basis of your answers to parts 1 and 2, what are the parametric equations representing the curtate cycloid?

Once the ant's head clears, he realizes that the bicyclist has made a turn, and is now traveling away from his home. So he drops off the bicycle tire and looks around. Fortunately, there is a set of train tracks nearby, headed back in the right direction. So the ant heads over to the train tracks to wait. After a while, a train goes by, heading in the right direction, and he manages to jump up and just catch the edge of the train wheel (without getting squished!).

The ant is still worried about getting dizzy, but the train wheel is slippery and has no spokes to climb, so he decides to just hang on to the edge of the wheel and hope for the best. Now, train wheels have a flange to keep the wheel running on the tracks. So, in this case, since the ant is hanging on to the very edge of the flange, the distance from the center of the wheel to the ant is actually greater than the radius of the wheel (Figure 7.1.14).

The setup here is essentially the same as when the ant climbed up the spoke on the bicycle wheel. We let b denote the distance from the center of the wheel to the ant, and we let t represent the angle the tire has rotated through. Additionally, we let $C = (x_C, y_C)$ represent the position of the center of the wheel and $A = (x_A, y_A)$ represent the position of the ant (Figure 7.1.14).

When the distance from the center of the wheel to the ant is greater than the radius of the wheel, his path of motion is called a prolate cycloid. A graph of a **prolate cycloid** is shown in the figure.

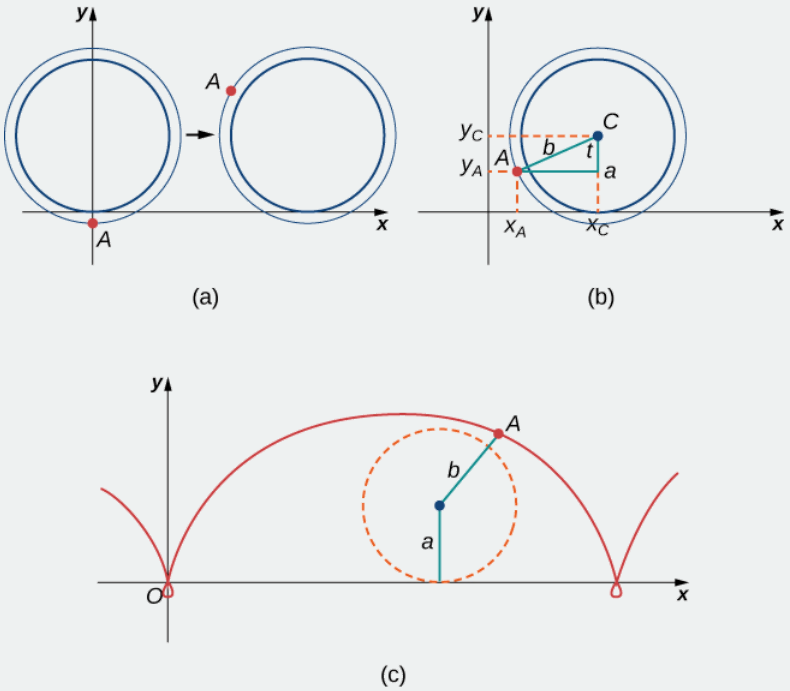


Figure 7.1.14: (a) The ant is hanging onto the flange of the train wheel. (b) The new setup, now that the ant has jumped onto the train wheel. (c) The ant travels along a prolate cycloid.

4. Using the same approach you used in parts 1– 3, find the parametric equations for the path of motion of the ant.
5. What do you notice about your answer to part 3 and your answer to part 4?

Notice that the ant is actually traveling backward at times (the "loops" in the graph), even though the train continues to move forward. He is probably going to be really dizzy by the time he gets home!

Key Concepts

- Parametric equations provide a convenient way to describe a curve. A parameter can represent time or some other meaningful quantity.
- It is often possible to eliminate the parameter in a parameterized curve to obtain a function or relation describing that curve.
- There is always more than one way to parameterize a curve.

- Parametric equations can describe complicated curves that are difficult or perhaps impossible to describe using rectangular coordinates.

Glossary

cycloid
the curve traced by a point on the rim of a circular wheel as the wheel rolls along a straight line without slippage

cusp
a pointed end or part where two curves meet

orientation
the direction that a point moves on a graph as the parameter increases

parameter
an independent variable that both x and y depend on in a parametric curve; usually represented by the variable t

parametric curve
the graph of the parametric equations $x(t)$ and $y(t)$ over an interval $a \leq t \leq b$ combined with the equations

parametric equations
the equations $x = x(t)$ and $y = y(t)$ that define a parametric curve

parameterization of a curve
rewriting the equation of a curve defined by a function $y = f(x)$ as parametric equations

---

This page titled 7.1: Parametric Equations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **11.1: Parametric Equations** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

7.2: Calculus of Parametric Curves

Learning Objectives

- Determine derivatives and equations of tangents for parametric curves.
- Find the area under a parametric curve.
- Use the equation for arc length of a parametric curve.
- Apply the formula for surface area to a volume generated by a parametric curve.

Now that we have introduced the concept of a parameterized curve, our next step is to learn how to work with this concept in the context of calculus. For example, if we know a parameterization of a given curve, is it possible to calculate the slope of a tangent line to the curve? How about the arc length of the curve? Or the area under the curve?

Another scenario: Suppose we would like to represent the location of a baseball after the ball leaves a pitcher's hand. If the position of the baseball is represented by the plane curve $(x(t), y(t))$ then we should be able to use calculus to find the speed of the ball at any given time. Furthermore, we should be able to calculate just how far that ball has traveled as a function of time.

Derivatives of Parametric Equations

We start by asking how to calculate the slope of a line tangent to a parametric curve at a point. Consider the plane curve defined by the parametric equations

$$x(t) = 2t + 3 \tag{7.2.1}$$
$$y(t) = 3t - 4 \tag{7.2.2}$$

within $-2 \leq t \leq 3$.

The graph of this curve appears in Figure 7.2.1. It is a line segment starting at $(-1, -10)$ and ending at $(9, 5)$.

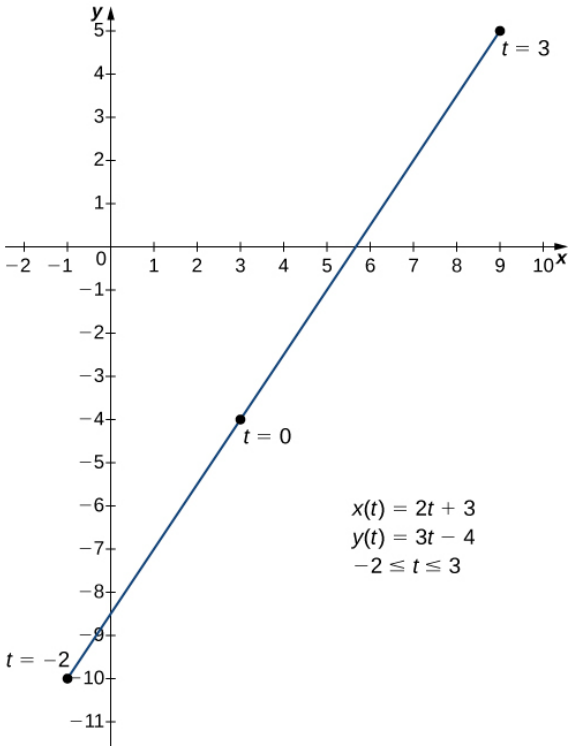


Figure 7.2.1: Graph of the line segment described by the given parametric equations.

We can eliminate the parameter by first solving Equation 7.2.1 for t:

$$x(t) = 2t + 3$$
$$x - 3 = 2t$$

$$t = \frac{x-3}{2}.$$

Substituting this into $y(t)$ (Equation 7.2.2), we obtain

$$y(t) = 3t - 4$$
$$y = 3\left(\frac{x-3}{2}\right) - 4$$
$$y = \frac{3x}{2} - \frac{9}{2} - 4$$
$$y = \frac{3x}{2} - \frac{17}{2}.$$

The slope of this line is given by $\frac{dy}{dx} = \frac{3}{2}$. Next we calculate $x'(t)$ and $y'(t)$. This gives $x'(t) = 2$ and $y'(t) = 3$. Notice that

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{3}{2}.$$

This is no coincidence, as outlined in the following theorem.

Derivative of Parametric Equations

Consider the plane curve defined by the parametric equations $x = x(t)$ and $y = y(t)$. Suppose that $x'(t)$ and $y'(t)$ exist, and assume that $x'(t) \neq 0$. Then the derivative $\frac{dy}{dx}$ is given by

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)}. \tag{7.2.3}$$

Proof

This theorem can be proven using the Chain Rule. In particular, assume that the parameter t can be eliminated, yielding a differentiable function $y = F(x)$. Then $y(t) = F(x(t))$. Differentiating both sides of this equation using the Chain Rule yields

$$y'(t) = F'\big(x(t)\big)x'(t),$$

so

$$F'\big(x(t)\big) = \frac{y'(t)}{x'(t)}.$$

But $F'\big(x(t)\big) = \frac{dy}{dx}$, which proves the theorem.

□

Equation 7.2.3 can be used to calculate derivatives of plane curves, as well as critical points. Recall that a critical point of a differentiable function $y = f(x)$ is any point $x = x_0$ such that either $f'(x_0) = 0$ or $f'(x_0)$ does not exist. Equation 7.2.3 gives a formula for the slope of a tangent line to a curve defined parametrically regardless of whether the curve can be described by a function $y = f(x)$ or not.

Example 7.2.1: Finding the Derivative of a Parametric Curve

Calculate the derivative $\frac{dy}{dx}$ for each of the following parametrically defined plane curves, and locate any critical points on their respective graphs.

a. $x(t) = t^2 - 3, \quad y(t) = 2t - 1, \quad \text{for } -3 \leq t \leq 4$
b. $x(t) = 2t + 1, \quad y(t) = t^3 - 3t + 4, \quad \text{for } -2 \leq t \leq 2$
c. $x(t) = 5\cos t, \quad y(t) = 5\sin t, \quad \text{for } 0 \leq t \leq 2\pi$

Solution

a. To apply Equation 7.2.3, first calculate $x'(t)$ and $y'(t)$:

$$x'(t) = 2t$$
$$y'(t) = 2.$$

Next substitute these into the equation:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$
$$\frac{dy}{dx} = \frac{2}{2t}$$
$$\frac{dy}{dx} = \frac{1}{t}.$$

This derivative is undefined when $t = 0$. Calculating $x(0)$ and $y(0)$ gives $x(0) = (0)^2 - 3 = -3$ and $y(0) = 2(0) - 1 = -1$, which corresponds to the point $(-3, -1)$ on the graph. The graph of this curve is a parabola opening to the right, and the point $(-3, -1)$ is its vertex as shown.

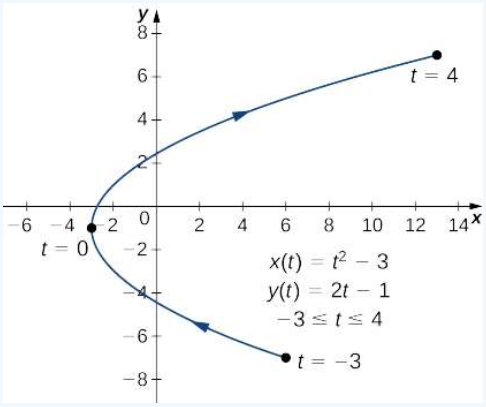


Figure 7.2.2: Graph of the parabola described by parametric equations in part a.

b. To apply Equation 7.2.3, first calculate $x'(t)$ and $y'(t)$:

$$x'(t) = 2$$
$$y'(t) = 3t^2 - 3\,.$$

Next substitute these into the equation:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$
$$\frac{dy}{dx} = \frac{3t^2 - 3}{2}\,.$$

This derivative is zero when $t = \pm 1$. When $t = -1$ we have

$$x(-1) = 2(-1) + 1 = -1 \text{ and } y(-1) = (-1)^3 - 3(-1) + 4 = -1 + 3 + 4 = 6\,,$$

which corresponds to the point $(-1, 6)$ on the graph. When $t = 1$ we have

$$x(1) = 2(1) + 1 = 3 \text{ and } y(1) = (1)^3 - 3(1) + 4 = 1 - 3 + 4 = 2,$$

which corresponds to the point $(3, 2)$ on the graph. The point $(3, 2)$ is a relative minimum and the point $(-1, 6)$ is a relative maximum, as seen in the following graph.

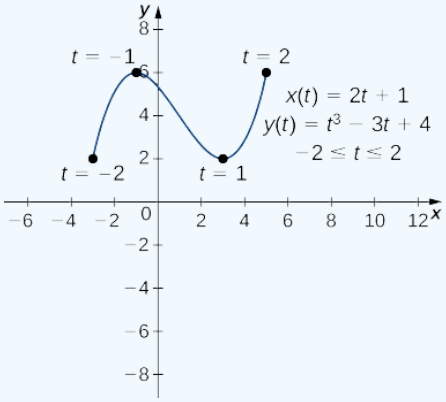


Figure 7.2.3: Graph of the curve described by parametric equations in part b.

c. To apply Equation 7.2.3, first calculate $x'(t)$ and $y'(t)$:

$$x'(t) = -5\sin t$$
$$y'(t) = 5\cos t.$$

Next substitute these into the equation:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$
$$\frac{dy}{dx} = \frac{5\cos t}{-5\sin t}$$
$$\frac{dy}{dx} = -\cot t.$$

This derivative is zero when $\cos t = 0$ and is undefined when $\sin t = 0$. This gives $t = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$, and 2π as critical points for t. Substituting each of these into $x(t)$ and $y(t)$, we obtain

t	$x(t)$	$y(t)$
0	5	0
$\frac{\pi}{2}$	0	5
π	−5	0
$\frac{3\pi}{2}$	0	−5
2π	5	0

These points correspond to the sides, top, and bottom of the circle that is represented by the parametric equations (Figure 7.2.4). On the left and right edges of the circle, the derivative is undefined, and on the top and bottom, the derivative equals zero.

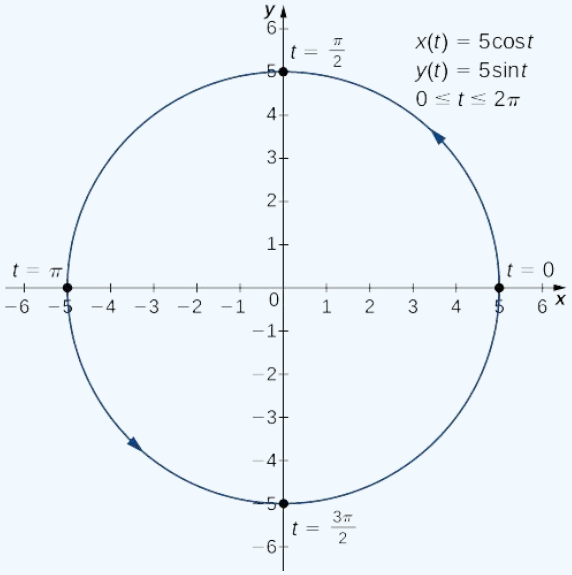


Figure 7.2.4: Graph of the curve described by parametric equations in part c.

Exercise 7.2.1

Calculate the derivative dy/dx for the plane curve defined by the equations

$$x(t) = t^2 - 4t, \quad y(t) = 2t^3 - 6t, \quad \text{for } -2 \le t \le 3$$

and locate any critical points on its graph.

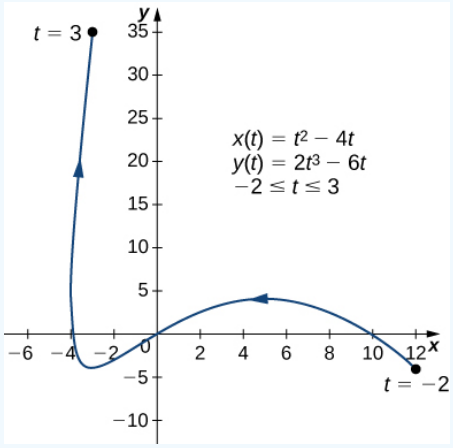
Hint

Calculate $x'(t)$ and $y'(t)$ and use Equation 7.2.3

Answer

$x'(t) = 2t - 4$ and $y'(t) = 6t^2 - 6$, so $\dfrac{dy}{dx} = \dfrac{6t^2 - 6}{2t - 4} = \dfrac{3t^2 - 3}{t - 2}$.

This expression is undefined when $t = 2$ and equal to zero when $t = \pm 1$.



Example 7.2.2: Finding a Tangent Line

Find the equation of the tangent line to the curve defined by the equations

$$x(t) = t^2 - 3, \quad y(t) = 2t - 1, \quad \text{for } -3 \le t \le 4$$

https://math.libretexts.org/@go/page/33834

when $t = 2$.

Solution

First find the slope of the tangent line using Equation 7.2.3, which means calculating $x'(t)$ and $y'(t)$:

$$x'(t) = 2t$$
$$y'(t) = 2.$$

Next substitute these into the equation:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$
$$\frac{dy}{dx} = \frac{2}{2t}$$
$$\frac{dy}{dx} = \frac{1}{t}.$$

When $t = 2$, $\frac{dy}{dx} = \frac{1}{2}$, so this is the slope of the tangent line. Calculating $x(2)$ and $y(2)$ gives

$$x(2) = (2)^2 - 3 = 1 \text{ and } y(2) = 2(2) - 1 = 3,$$

which corresponds to the point $(1, 3)$ on the graph (Figure 7.2.5). Now use the point-slope form of the equation of a line to find the equation of the tangent line:

$$y - y_0 = m(x - x_0)$$
$$y - 3 = \frac{1}{2}(x - 1)$$
$$y - 3 = \frac{1}{2}x - \frac{1}{2}$$
$$y = \frac{1}{2}x + \frac{5}{2}.$$

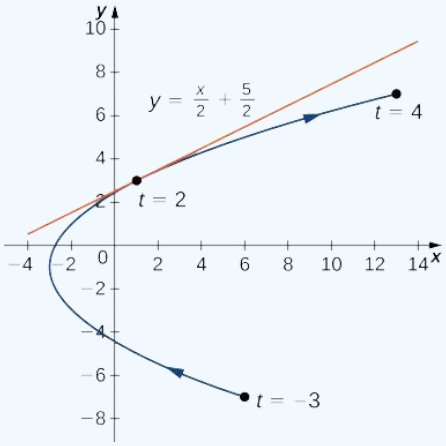


Figure 7.2.5: Tangent line to the parabola described by the given parametric equations when $t = 2$.

? Exercise 7.2.2

Find the equation of the tangent line to the curve defined by the equations

$$x(t) = t^2 - 4t, \quad y(t) = 2t^3 - 6t, \quad \text{for } -2 \le t \le 6 \quad \text{when } t = 5.$$

Hint

Calculate $x'(t)$ and $y'(t)$ and use Equation 7.2.3

Answer

The equation of the tangent line is $y = 24x + 100$.

Second-Order Derivatives

Our next goal is to see how to take the second derivative of a function defined parametrically. The second derivative of a function $y = f(x)$ is defined to be the derivative of the first derivative; that is,

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dx}\right]. \tag{7.2.4}$$

Since

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt},$$

we can replace the y on both sides of Equation 7.2.4 with $\frac{dy}{dx}$. This gives us

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{(d/dt)(dy/dx)}{dx/dt}. \tag{7.2.5}$$

If we know dy/dx as a function of t, then this formula is straightforward to apply

Example 7.2.3: Finding a Second Derivative

Calculate the second derivative d^2y/dx^2 for the plane curve defined by the parametric equations $x(t) = t^2 - 3, \quad y(t) = 2t - 1, \quad \text{for } -3 \le t \le 4.$

Solution

From Example 7.2.1 we know that $\frac{dy}{dx} = \frac{2}{2t} = \frac{1}{t}$. Using Equation 7.2.5, we obtain

$$\frac{d^2y}{dx^2} = \frac{(d/dt)(dy/dx)}{dx/dt} = \frac{(d/dt)(1/t)}{2t} = \frac{-t^{-2}}{2t} = -\frac{1}{2t^3}.$$

Exercise 7.2.3

Calculate the second derivative d^2y/dx^2 for the plane curve defined by the equations

$$x(t) = t^2 - 4t, \quad y(t) = 2t^3 - 6t, \quad \text{for } -2 \le t \le 3$$

and locate any critical points on its graph.

Hint

Start with the solution from the previous exercise, and use Equation 7.2.5

Answer

$\frac{d^2y}{dx^2} = \frac{3t^2 - 12t + 3}{2(t-2)^3}$. Critical points $(5, 4)$, $(-3, -4)$,and $(-4, 6)$.

Integrals Involving Parametric Equations

Now that we have seen how to calculate the derivative of a plane curve, the next question is this: How do we find the area under a curve defined parametrically? Recall the cycloid defined by these parametric equations

$$\begin{aligned} x(t) &= t - \sin t \\ y(t) &= 1 - \cos t. \end{aligned}$$

Suppose we want to find the area of the shaded region in the following graph.

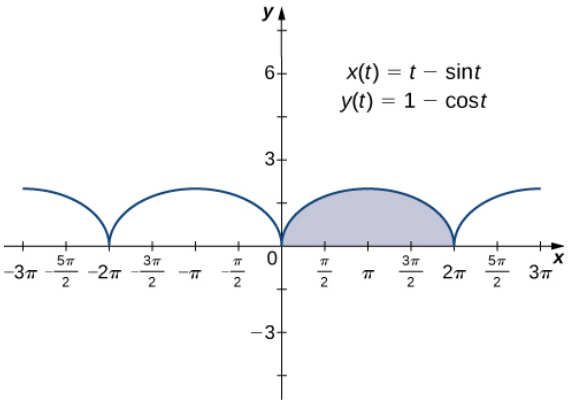


Figure 7.2.6: Graph of a cycloid with the arch over $[0, 2\pi]$ highlighted.

To derive a formula for the area under the curve defined by the functions

$$x = x(t)$$
$$y = y(t)$$

where $a \le t \le b$.

We assume that $x(t)$ is differentiable and start with an equal partition of the interval $a \le t \le b$. Suppose $t_0 = a < t_1 < t_2 < \cdots < t_n = b$ and consider the following graph.

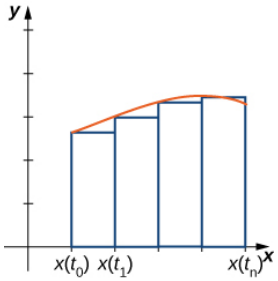


Figure 7.2.7: Approximating the area under a parametrically defined curve.

We use rectangles to approximate the area under the curve. The height of a typical rectangle in this parametrization is $y(x(\bar{t}_i))$ for some value $\bar{t}_i$ in the ith subinterval, and the width can be calculated as $x(t_i) - x(t_{i-1})$. Thus the area of the ith rectangle is given by

$$A_i = y(x(\bar{t}_i))(x(t_i) - x(t_{i-1})).$$

Then a Riemann sum for the area is

$$A_n = \sum_{i=1}^{n} y(x(\bar{t}_i))(x(t_i) - x(t_{i-1})).$$

Multiplying and dividing each area by $t_i - t_{i-1}$ gives

$$\begin{aligned} A_n &= \sum_{i=1}^{n} y(x(\bar{t}_i)) \left(\frac{x(t_i) - x(t_{i-1})}{t_i - t_{i-1}} \right) (t_i - t_{i-1}) \\ &= \sum_{i=1}^{n} y(x(\bar{t}_i)) \left(\frac{x(t_i) - x(t_{i-1})}{\Delta t} \right) \Delta t. \end{aligned}$$

Taking the limit as n approaches infinity gives

$$A = \lim_{n \to \infty} A_n = \int_a^b y(t) x'(t)\, dt.$$

This leads to the following theorem.

Area under a Parametric Curve

Consider the non-self-intersecting plane curve defined by the parametric equations

$$x = x(t), \quad y = y(t), \quad \text{for } a \le t \le b$$

and assume that $x(t)$ is differentiable. The area under this curve is given by

$$A = \int_a^b y(t)x'(t)\,dt. \tag{7.2.6}$$

Example 7.2.4: Finding the Area under a Parametric Curve

Find the area under the curve of the cycloid defined by the equations

$$x(t) = t - \sin t, \quad y(t) = 1 - \cos t, \quad \text{for } 0 \le t \le 2\pi.$$

Solution

Using Equation 7.2.6, we have

$$\begin{aligned} A &= \int_a^b y(t)x'(t)\,dt \\ &= \int_0^{2\pi} (1-\cos t)(1-\cos t)\,dt \\ &= \int_0^{2\pi} (1 - 2\cos t + \cos^2 t)\,dt \\ &= \int_0^{2\pi} \left(1 - 2\cos t + \frac{1+\cos(2t)}{2}\right) dt \\ &= \int_0^{2\pi} \left(\frac{3}{2} - 2\cos t + \frac{\cos(2t)}{2}\right) dt \\ &= \frac{3t}{2} - 2\sin t + \frac{\sin(2t)}{4}\Big|_0^{2\pi} \\ &= 3\pi \end{aligned}$$

Exercise 7.2.4

Find the area under the curve of the hypocycloid defined by the equations

$$x(t) = 3\cos t + \cos(3t), \quad y(t) = 3\sin t - \sin(3t), \quad \text{for } 0 \le t \le \pi.$$

Hint

Use Equation 7.2.6, along with the identities $\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)]$ and $\sin^2 t = \frac{1-\cos(2t)}{2}$.

Answer

$A = 3\pi$ (Note that the integral formula actually yields a negative answer. This is due to the fact that $x(t)$ is a decreasing function over the interval $[0, \pi]$; that is, the curve is traced from right to left.)

Arc Length of a Parametric Curve

In addition to finding the area under a parametric curve, we sometimes need to find the arc length of a parametric curve. In the case of a line segment, arc length is the same as the distance between the endpoints. If a particle travels from point A to point B along a curve, then the distance that particle travels is the arc length. To develop a formula for arc length, we start with an approximation by line segments as shown in the following graph.

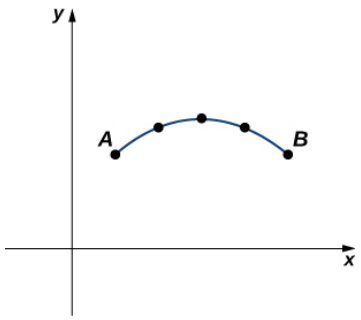


Figure 7.2.7: Approximation of a curve by line segments.

Given a plane curve defined by the functions $x = x(t), \quad y = y(t), \quad$ for $a \le t \le b$, we start by partitioning the interval $[a, b]$ into n equal subintervals: $t_0 = a < t_1 < t_2 < \cdots < t_n = b$. The width of each subinterval is given by $\Delta t = (b-a)/n$. We can calculate the length of each line segment:

$$d_1 = \sqrt{(x(t_1) - x(t_0))^2 + (y(t_1) - y(t_0))^2}$$

$$d_2 = \sqrt{(x(t_2) - x(t_1))^2 + (y(t_2) - y(t_1))^2}$$

etc.

Then add these up. We let s denote the exact arc length and s_n denote the approximation by n line segments:

$$s \approx \sum_{k=1}^{n} s_k = \sum_{k=1}^{n} \sqrt{(x(t_k) - x(t_{k-1}))^2 + (y(t_k) - y(t_{k-1}))^2} \; . \tag{7.2.7}$$

If we assume that $x(t)$ and $y(t)$ are differentiable functions of t, then the Mean Value Theorem applies, so in each subinterval $[t_{k-1}, t_k]$ there exist $\hat{t_k}$ and $\tilde{t_k}$ such that

$$x(t_k) - x(t_{k-1}) = x'(\hat{t_k})(t_k - t_{k-1}) = x'(\hat{t_k})\,\Delta t$$

$$y(t_k) - y(t_{k-1}) = y'(\tilde{t_k})(t_k - t_{k-1}) = y'(\tilde{t_k})\,\Delta t.$$

Therefore Equation 7.2.7 becomes

$$\begin{aligned} s \approx \sum_{k=1}^{n} s_k &= \sum_{k=1}^{n} \sqrt{(x'(\hat{t_k})\Delta t)^2 + (y'(\tilde{t_k})\Delta t)^2} \\ &= \sum_{k=1}^{n} \sqrt{(x'(\hat{t_k}))^2(\Delta t)^2 + (y'(\tilde{t_k}))^2(\Delta t)^2} \\ &= \sum_{k=1}^{n} \sqrt{(x'(\hat{t_k}))^2 + (y'(\tilde{t_k}))^2}\,\Delta t. \end{aligned}$$

This is a Riemann sum that approximates the arc length over a partition of the interval $[a, b]$. If we further assume that the derivatives are continuous and let the number of points in the partition increase without bound, the approximation approaches the exact arc length. This gives

$$\begin{aligned} s &= \lim_{n\to\infty} \sum_{k=1}^{n} s_k \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \sqrt{(x'(\hat{t_k}))^2 + (y'(\tilde{t_k}))^2}\,\Delta t \\ &= \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2}\,dt. \end{aligned}$$

When taking the limit, the values of $\hat{t_k}$ and $\tilde{t_k}$ are both contained within the same ever-shrinking interval of width Δt, so they must converge to the same value.

We can summarize this method in the following theorem.

Arc Length of a Parametric Curve

Consider the plane curve defined by the parametric equations

$$x = x(t), \quad y = y(t), \quad \text{for } t_1 \le t \le t_2$$

and assume that $x(t)$ and $y(t)$ are differentiable functions of t. Then the arc length of this curve is given by

$$s = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \tag{7.2.8}$$

At this point a side derivation leads to a previous formula for arc length. In particular, suppose the parameter can be eliminated, leading to a function $y = F(x)$. Then $y(t) = F(x(t))$ and the Chain Rule gives

$$y'(t) = F'\big(x(t)\big)x'(t).$$

Substituting this into Equation 7.2.8 gives

$$\begin{aligned} s &= \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(F'(x)\frac{dx}{dt}\right)^2}\, dt \\ &= \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 \left(1 + (F'(x))^2\right)}\, dt \\ &= \int_{t_1}^{t_2} x'(t)\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dt. \end{aligned}$$

Here we have assumed that $x'(t) > 0$, which is a reasonable assumption. The Chain Rule gives $dx = x'(t)\, dt$, and letting $a = x(t_1)$ and $b = x(t_2)$ we obtain the formula

$$s = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx,$$

which is the formula for arc length obtained in the Introduction to the Applications of Integration.

Example 7.2.5: Finding the Arc Length of a Parametric Curve

Find the arc length of the semicircle defined by the equations

$$x(t) = 3\cos t, \quad y(t) = 3\sin t, \quad \text{for } 0 \le t \le \pi.$$

Solution

The values $t = 0$ to $t = \pi$ trace out the blue curve in Figure 7.2.8. To determine its length, use Equation 7.2.8:

$$\begin{aligned} s &= \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \\ &= \int_0^{\pi} \sqrt{(-3\sin t)^2 + (3\cos t)^2}\, dt \\ &= \int_0^{\pi} \sqrt{9\sin^2 t + 9\cos^2 t}\, dt \\ &= \int_0^{\pi} \sqrt{9(\sin^2 t + \cos^2 t)}\, dt \\ &= \int_0^{\pi} 3\, dt = 3t\Big|_0^{\pi} \\ &= 3\pi \text{ units}. \end{aligned}$$

Note that the formula for the arc length of a semicircle is πr and the radius of this circle is 3. This is a great example of using calculus to derive a known formula of a geometric quantity.

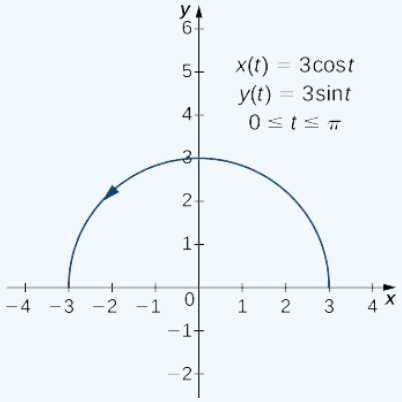


Figure 7.2.8: The arc length of the semicircle is equal to its radius times π.

Exercise 7.2.5

Find the arc length of the curve defined by the equations

$$x(t) = 3t^2, \quad y(t) = 2t^3, \quad \text{for } 1 \le t \le 3.$$

Hint

Use Equation 7.2.8

Answer

$s = 2(10^{3/2} - 2^{3/2}) \approx 57.589$ units

We now return to the problem posed at the beginning of the section about a baseball leaving a pitcher's hand. Ignoring the effect of air resistance (unless it is a curve ball!), the ball travels a parabolic path. Assuming the pitcher's hand is at the origin and the ball travels left to right in the direction of the positive x-axis, the parametric equations for this curve can be written as

$$x(t) = 140t, \quad y(t) = -16t^2 + 2t$$

where t represents time. We first calculate the distance the ball travels as a function of time. This distance is represented by the arc length. We can modify the arc length formula slightly. First rewrite the functions $x(t)$ and $y(t)$ using v as an independent variable, so as to eliminate any confusion with the parameter t:

$$x(v) = 140v, \quad y(v) = -16v^2 + 2v.$$

Then we write the arc length formula as follows:

$$\begin{aligned} s(t) &= \int_0^t \sqrt{(\frac{dx}{dv})^2 + (\frac{dy}{dv})^2}\, dv \\ &= \int_0^t \sqrt{140^2 + (-32v+2)^2}\, dv \end{aligned}$$

The variable v acts as a dummy variable that disappears after integration, leaving the arc length as a function of time t. To integrate this expression we can use a formula from Appendix A,

$$\int \sqrt{a^2+u^2}\, du = \frac{u}{2}\sqrt{a^2+u^2} + \frac{a^2}{2}\ln|\, u + \sqrt{a^2+u^2}\,| + C.$$

We set $a = 140$ and $u = -32v + 2$. This gives $du = -32\, dv$, so $dv = -\dfrac{1}{32}\, du$. Therefore

$$\begin{aligned} \int \sqrt{140^2 + (-32v+2)^2}\, dv &= -\frac{1}{32}\int \sqrt{a^2+u^2}\, du \\ &= -\frac{1}{32}\left[\frac{(-32v+2)}{2}\sqrt{140^2+(-32v+2)^2} + \frac{140^2}{2}\ln|\,(-32v+2) + \sqrt{140^2+(-32v+2)^2}\,| + C\right] \end{aligned}$$

and

$$\begin{aligned} s(t) &= -\frac{1}{32}\left[\frac{(-32t+2)}{2}\sqrt{140^2+(-32t+2)^2}+\frac{140^2}{2}\ln\left|(-32t+2)+\sqrt{140^2+(-32t+2)^2}\right|\right] \\ &\quad +\frac{1}{32}\left[\sqrt{140^2+2^2}+\frac{140^2}{2}\ln\left|2+\sqrt{140^2+2^2}\right|\right] \\ &= \left(\frac{t}{2}-\frac{1}{32}\right)\sqrt{1024t^2-128t+19604}-\frac{1225}{4}\ln\left|(-32t+2)+\sqrt{1024t^2-128t+19604}\right|+\frac{\sqrt{19604}}{32}+\frac{1225}{4}\ln \\ &\quad (2+\sqrt{19604}) \end{aligned}$$

This function represents the distance traveled by the ball as a function of time. To calculate the speed, take the derivative of this function with respect to t. While this may seem like a daunting task, it is possible to obtain the answer directly from the Fundamental Theorem of Calculus:

$$\frac{d}{dx}\int_a^x f(u)\,du = f(x).$$

Therefore

$$\begin{aligned} s'(t) &= \frac{d}{dt}\left[s(t)\right] \\ &= \frac{d}{dt}\left[\int_0^t \sqrt{140^2+(-32v+2)^2}\,dv\right] \\ &= \sqrt{140^2+(-32t+2)^2} \\ &= \sqrt{1024t^2-128t+19604} \\ &= 2\sqrt{256t^2-32t+4901}. \end{aligned}$$

One third of a second after the ball leaves the pitcher's hand, the distance it travels is equal to

$$\begin{aligned} s\left(\frac{1}{3}\right) &= \left(\frac{1/3}{2}-\frac{1}{32}\right)\sqrt{1024\left(\frac{1}{3}\right)^2-128\left(\frac{1}{3}\right)+19604} \\ &\quad -\frac{1225}{4}\ln\left|\left(-32\left(\frac{1}{3}\right)+2\right)+\sqrt{1024\left(\frac{1}{3}\right)^2-128\left(\frac{1}{3}\right)+19604}\right| \\ &\quad +\frac{\sqrt{19604}}{32}+\frac{1225}{4}\ln(2+\sqrt{19604}) \\ &\approx 46.69 \text{ feet}. \end{aligned}$$

This value is just over three quarters of the way to home plate. The speed of the ball is

$$s'\left(\tfrac{1}{3}\right) = 2\sqrt{256\left(\tfrac{1}{3}\right)^2-32\left(\tfrac{1}{3}\right)+4901} \approx 140.27 \text{ ft/s}.$$

This speed translates to approximately 95 mph—a major-league fastball.

Surface Area Generated by a Parametric Curve

Recall the problem of finding the surface area of a volume of revolution. In Curve Length and Surface Area, we derived a formula for finding the surface area of a volume generated by a function $y=f(x)$ from $x=a$ to $x=b$, revolved around the x-axis:

$$S = 2\pi\int_a^b f(x)\sqrt{1+(f'(x))^2}\,dx.$$

We now consider a volume of revolution generated by revolving a parametrically defined curve $x=x(t), \quad y=y(t), \quad$ for $a\le t\le b$ around the x-axis as shown in Figure 7.2.9.

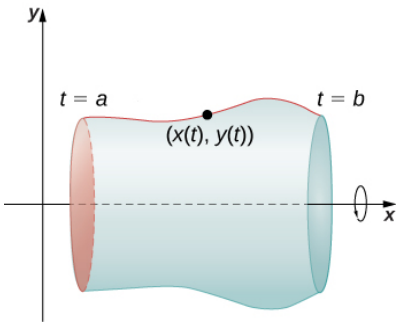


Figure 7.2.9: A surface of revolution generated by a parametrically defined curve.

The analogous formula for a parametrically defined curve is

$$S = 2\pi \int_a^b y(t)\sqrt{(x'(t))^2 + (y'(t))^2}\, dt \tag{7.2.9}$$

provided that $y(t)$ is not negative on $[a, b]$.

Example 7.2.6: Finding Surface Area

Find the surface area of a sphere of radius r centered at the origin.

Solution

We start with the curve defined by the equations

$$x(t) = r\cos t, \quad y(t) = r\sin t, \quad \text{for } 0 \le t \le \pi.$$

This generates an upper semicircle of radius r centered at the origin as shown in the following graph.

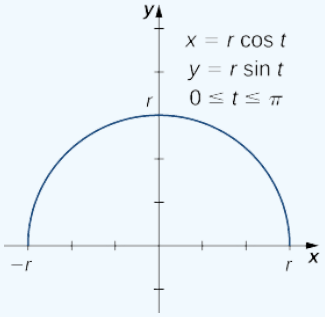


Figure 7.2.10: A semicircle generated by parametric equations.

When this curve is revolved around the x-axis, it generates a sphere of radius r. To calculate the surface area of the sphere, we use Equation 7.2.9:

$$
\begin{aligned}
S &= 2\pi \int_a^b y(t)\sqrt{(x'(t))^2 + (y'(t))^2}\, dt \\
&= 2\pi \int_0^{\pi} r\sin t \sqrt{(-r\sin t)^2 + (r\cos t)^2}\, dt \\
&= 2\pi \int_0^{\pi} r\sin t \sqrt{r^2\sin^2 t + r^2\cos^2 t}\, dt \\
&= 2\pi \int_0^{\pi} r\sin t \sqrt{r^2(\sin^2 t + \cos^2 t)}\, dt \\
&= 2\pi \int_0^{\pi} r^2 \sin t\, dt \\
&= 2\pi r^2 \left(-\cos t\Big|_0^{\pi}\right) \\
&= 2\pi r^2(-\cos\pi + \cos 0) \\
&= 4\pi r^2 \text{ units}^2.
\end{aligned}
$$

This is, in fact, the formula for the surface area of a sphere.

? Exercise 7.2.6

Find the surface area generated when the plane curve defined by the equations

$$x(t) = t^3, \quad y(t) = t^2, \quad \text{for } 0 \le t \le 1$$

is revolved around the x-axis.

Hint

Use Equation 7.2.9 When evaluating the integral, use a u-substitution.

Answer

$$A = \frac{\pi(494\sqrt{13} + 128)}{1215} \text{ units}^2$$

Key Concepts

- The derivative of the parametrically defined curve $x = x(t)$ and $y = y(t)$ can be calculated using the formula $\frac{dy}{dx} = \frac{y'(t)}{x'(t)}$. Using the derivative, we can find the equation of a tangent line to a parametric curve.
- The area between a parametric curve and the x-axis can be determined by using the formula $A = \int_{t_1}^{t_2} y(t)x'(t)\, dt$.
- The arc length of a parametric curve can be calculated by using the formula

$$s = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

- The surface area of a volume of revolution revolved around the x-axis is given by

$$S = 2\pi \int_a^b y(t)\sqrt{(x'(t))^2 + (y'(t))^2}\, dt.$$

- If the curve is revolved around the y-axis, then the formula is

$$S = 2\pi \int_a^b x(t)\sqrt{(x'(t))^2 + (y'(t))^2}\, dt.$$

Key Equations

- **Derivative of parametric equations**

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)}$$

- **Second-order derivative of parametric equations**

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

- **Area under a parametric curve**

$$A = \int_a^b y(t)x'(t)\,dt$$

- **Arc length of a parametric curve**

$$s = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt$$

- **Surface area generated by a parametric curve**

$$S = 2\pi \int_a^b y(t)\sqrt{(x'(t))^2 + (y'(t))^2}\,dt$$

This page titled 7.2: Calculus of Parametric Curves is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **11.2: Calculus of Parametric Curves** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

7.3: Polar Coordinates

Learning Objectives

- Locate points in a plane by using polar coordinates.
- Convert points between rectangular and polar coordinates.
- Sketch polar curves from given equations.
- Convert equations between rectangular and polar coordinates.
- Identify symmetry in polar curves and equations.

The rectangular coordinate system (or Cartesian plane) provides a means of mapping points to ordered pairs and ordered pairs to points. This is called a *one-to-one mapping* from points in the plane to ordered pairs. The polar coordinate system provides an alternative method of mapping points to ordered pairs. In this section we see that in some circumstances, polar coordinates can be more useful than rectangular coordinates.

Defining Polar Coordinates

To find the coordinates of a point in the polar coordinate system, consider Figure 7.3.1. The point P has Cartesian coordinates (x, y). The line segment connecting the origin to the point P measures the distance from the origin to P and has length r. The angle between the positive x-axis and the line segment has measure θ. This observation suggests a natural correspondence between the coordinate pair (x, y) and the values r and θ. This correspondence is the basis of the **polar coordinate system**. Note that every point in the Cartesian plane has two values (hence the term ordered pair) associated with it. In the polar coordinate system, each point also has two values associated with it: r and θ.

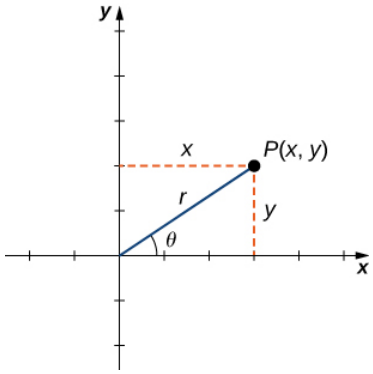


Figure 7.3.1: An arbitrary point in the Cartesian plane.

Using right-triangle trigonometry, the following equations are true for the point P:

$$\cos\theta = \frac{x}{r} \text{ so } x = r\cos\theta$$

$$\sin\theta = \frac{y}{r} \text{ so } y = r\sin\theta.$$

Furthermore,

$$r^2 = x^2 + y^2$$

and

$$\tan\theta = \frac{y}{x}.$$

Each point (x, y) in the Cartesian coordinate system can therefore be represented as an ordered pair (r, θ) in the polar coordinate system. The first coordinate is called the **radial coordinate** and the second coordinate is called the **angular coordinate**. Every point in the plane can be represented in this form.

Note that the equation $\tan\theta = y/x$ has an infinite number of solutions for any ordered pair (x, y). However, if we restrict the solutions to values between 0 and 2π then we can assign a unique solution to the quadrant in which the original point (x, y) is

located. Then the corresponding value of r is positive, so $r^2 = x^2 + y^2$.

Converting Points between Coordinate Systems

Given a point P in the plane with Cartesian coordinates (x, y) and polar coordinates (r, θ), the following conversion formulas hold true:

$$x = r\cos\theta \tag{7.3.1}$$

$$y = r\sin\theta \tag{7.3.2}$$

and

$$r^2 = x^2 + y^2 \tag{7.3.3}$$

$$\tan\theta = \frac{y}{x} \tag{7.3.4}$$

These formulas can be used to convert from rectangular to polar or from polar to rectangular coordinates. Notice that Equation 7.3.3 is the *Pythagorean theorem*. (Figure 7.3.1).

Example 7.3.1: Converting between Rectangular and Polar Coordinates

Convert each of the following points into polar coordinates.

a. $(1, 1)$
b. $(-3, 4)$
c. $(0, 3)$
d. $(5\sqrt{3}, -5)$

Convert each of the following points into rectangular coordinates.

e. $(3, \pi/3)$
f. $(2, 3\pi/2)$
g. $(6, -5\pi/6)$

Solution

a. Use $x = 1$ and $y = 1$ in Equation 7.3.3:

$$\begin{aligned} r^2 &= x^2 + y^2 \\ &= 1^2 + 1^2 \\ r &= \sqrt{2} \end{aligned}$$

and via Equation 7.3.4

$$\begin{aligned} \tan\theta &= \frac{y}{x} = \frac{1}{1} = 1 \\ \theta &= \frac{\pi}{4}. \end{aligned}$$

Therefore this point can be represented as $(\sqrt{2}, \frac{\pi}{4})$ in polar coordinates.

b. Use $x = -3$ and $y = 4$ in Equation 7.3.3:

$$\begin{aligned} r^2 &= x^2 + y^2 = (-3)^2 + (4)^2 \\ r &= 5 \end{aligned}$$

and via Equation 7.3.4

$$\tan\theta = \frac{y}{x} = -\frac{4}{3}$$

$$\theta = \arctan(-\frac{4}{3}) + \pi \approx 2.21.$$

Therefore this point can be represented as $(5, 2.21)$ in polar coordinates.

c. Use $x = 0$ and $y = 3$ in Equation 7.3.3:

$$r^2 = x^2 + y^2 = (3)^2 + (0)^2 = 9 + 0 \quad r = 3$$

and via Equation 7.3.4

$$\tan\theta = \frac{y}{x} = \frac{3}{0}\ .$$

Direct application of the second equation leads to division by zero. Graphing the point $(0, 3)$ on the rectangular coordinate system reveals that the point is located on the positive y-axis. The angle between the positive x-axis and the positive y-axis is $\frac{\pi}{2}$. Therefore this point can be represented as $(3, \frac{\pi}{2})$ in polar coordinates.

d. Use $x = 5\sqrt{3}$ and $y = -5$ in Equation 7.3.3:

$$r^2 = x^2 + y^2 = (5\sqrt{3})^2 + (-5)^2 = 75 + 25$$

$$r = 10$$

and via Equation 7.3.4

$$\tan\theta = \frac{y}{x} = \frac{-5}{5\sqrt{3}} = -\frac{\sqrt{3}}{3}$$

$$\theta = -\frac{\pi}{6}\ .$$

Therefore this point can be represented as $(10, -\frac{\pi}{6})$ in polar coordinates.

e. Use $r = 3$ and $\theta = \frac{\pi}{3}$ in Equation 7.3.1:

$$x = r\cos\theta = 3\cos(\frac{\pi}{3}) = 3(\frac{1}{2}) = \frac{3}{2}$$

and

$$y = r\sin\theta = 3\sin(\frac{\pi}{3}) = 3(\frac{\sqrt{3}}{2}) = \frac{3\sqrt{3}}{2}\ .$$

Therefore this point can be represented as $(\frac{3}{2}, \frac{3\sqrt{3}}{2})$ in rectangular coordinates.

f. Use $r = 2$ and $\theta = \frac{3\pi}{2}$ in Equation 7.3.1:

$$x = r\cos\theta = 2\cos(\frac{3\pi}{2}) = 2(0) = 0$$

and

$$y = r\sin\theta = 2\sin(\frac{3\pi}{2}) = 2(-1) = -2.$$

Therefore this point can be represented as $(0, -2)$ in rectangular coordinates.

g. Use $r = 6$ and $\theta = -\frac{5\pi}{6}$ in Equation 7.3.1:

$$x = r\cos\theta = 6\cos(-\frac{5\pi}{6}) = 6(-\frac{\sqrt{3}}{2}) = -3\sqrt{3}$$

and

$$y = r\sin\theta = 6\sin(-\frac{5\pi}{6}) = 6(-\frac{1}{2}) = -3\ .$$

Therefore this point can be represented as $(-3\sqrt{3}, -3)$ in rectangular coordinates.

Exercise 7.3.1

Convert $(-8, -8)$ into polar coordinates and $(4, \frac{2\pi}{3})$ into rectangular coordinates.

Hint

Use Equation 7.3.3and Equation 7.3.1. Make sure to check the quadrant when calculating θ.

Answer

$(8\sqrt{2}, \frac{5\pi}{4})$ and $(-2, 2\sqrt{3})$

The polar representation of a point is not unique. For example, the polar coordinates $(2, \frac{\pi}{3})$ and $(2, \frac{7\pi}{3})$ both represent the point $(1, \sqrt{3})$ in the rectangular system. Also, the value of r can be negative. Therefore, the point with polar coordinates $(-2, \frac{4\pi}{3})$ also represents the point $(1, \sqrt{3})$ in the rectangular system, as we can see by using Equation 7.3.1:

$$x = r\cos\theta = -2\cos(\frac{4\pi}{3}) = -2(-\frac{1}{2}) = 1$$

and

$$y = r\sin\theta = -2\sin(\frac{4\pi}{3}) = -2(-\frac{\sqrt{3}}{2}) = \sqrt{3}.$$

Every point in the plane has an infinite number of representations in polar coordinates. However, each point in the plane has only one representation in the rectangular coordinate system.

Note that the polar representation of a point in the plane also has a visual interpretation. In particular, r is the directed distance that the point lies from the origin, and θ measures the angle that the line segment from the origin to the point makes with the positive x-axis. Positive angles are measured in a counterclockwise direction and negative angles are measured in a clockwise direction. The polar coordinate system appears in Figure 7.3.2.

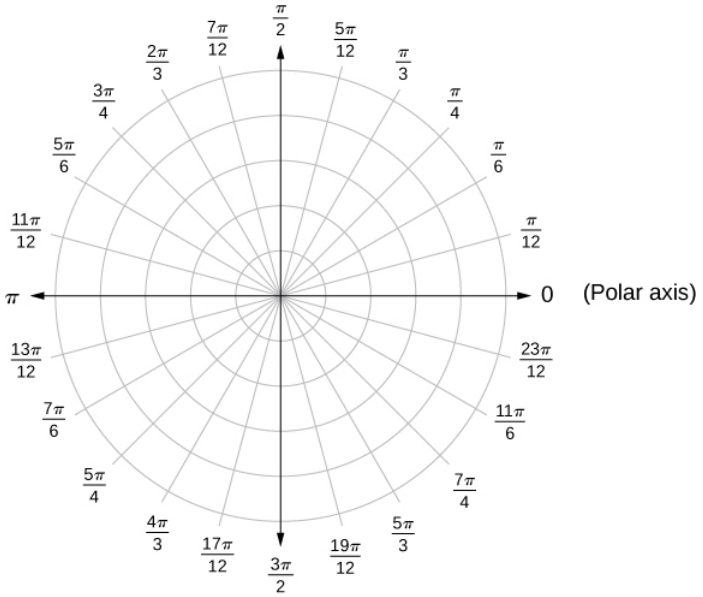


Figure 7.3.2: The polar coordinate system.

The line segment starting from the center of the graph going to the right (called the positive x-axis in the Cartesian system) is the **polar axis.** The center point is the **pole**, or origin, of the coordinate system, and corresponds to $r = 0$. The innermost circle shown in Figure 7.3.2 contains all points a distance of 1 unit from the pole, and is represented by the equation $r = 1$. Then $r = 2$ is the set

of points 2 units from the pole, and so on. The line segments emanating from the pole correspond to fixed angles. To plot a point in the polar coordinate system, start with the angle. If the angle is positive, then measure the angle from the polar axis in a counterclockwise direction. If it is negative, then measure it clockwise. If the value of r is positive, move that distance along the terminal ray of the angle. If it is negative, move along the ray that is opposite the terminal ray of the given angle.

Example 7.3.2: Plotting Points in the Polar Plane

Plot each of the following points on the polar plane.

a. $(2, \frac{\pi}{4})$

b. $(-3, \frac{2\pi}{3})$

c. $(4, \frac{5\pi}{4})$

Solution

The three points are plotted in Figure 7.3.3.

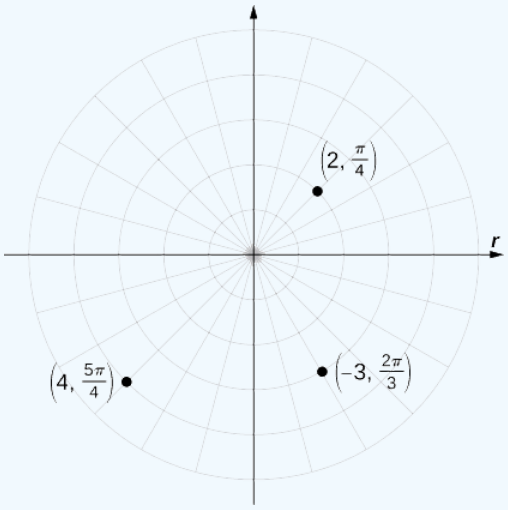


Figure 7.3.3: Three points plotted in the polar coordinate system.

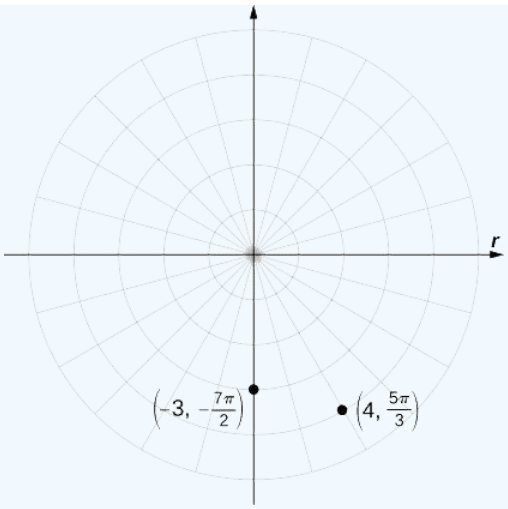
Exercise 7.3.2

Plot $(4, \frac{5\pi}{3})$ and $(-3, -\frac{7\pi}{2})$ on the polar plane.

Hint

Start with θ, then use r.

Answer



Polar Curves

Now that we know how to plot points in the polar coordinate system, we can discuss how to plot curves. In the rectangular coordinate system, we can graph a function $y = f(x)$ and create a curve in the Cartesian plane. In a similar fashion, we can graph a curve that is generated by a function $r = f(\theta)$.

The general idea behind graphing a function in polar coordinates is the same as graphing a function in rectangular coordinates. Start with a list of values for the independent variable (θ in this case) and calculate the corresponding values of the dependent variable r. This process generates a list of ordered pairs, which can be plotted in the polar coordinate system. Finally, connect the points, and take advantage of any patterns that may appear. The function may be periodic, for example, which indicates that only a limited number of values for the independent variable are needed.

Problem-Solving Strategy: Plotting a Curve in Polar Coordinates

1. Create a table with two columns. The first column is for θ, and the second column is for r.
2. Create a list of values for θ.
3. Calculate the corresponding r values for each θ.
4. Plot each ordered pair (r, θ) on the coordinate axes.
5. Connect the points and look for a pattern.

Example 7.3.3: Graphing a Function in Polar Coordinates

Graph the curve defined by the function $r = 4\sin\theta$. Identify the curve and rewrite the equation in rectangular coordinates.

Solution

Because the function is a multiple of a sine function, it is periodic with period 2π, so use values for θ between 0 and 2π. The result of steps 1–3 appear in the following table. Figure $7.3.4$ shows the graph based on this table.

Table 7.3.1: Table of values for this polar function.

θ	$r = 4\sin\theta$	θ	$r = 4\sin\theta$
0	0	π	0
$\frac{\pi}{6}$	2	$\frac{7\pi}{6}$	-2
$\frac{\pi}{4}$	$2\sqrt{2} \approx 2.8$	$\frac{5\pi}{4}$	$-2\sqrt{2} \approx -2.8$
$\frac{\pi}{3}$	$2\sqrt{3} \approx 3.4$	$\frac{4\pi}{3}$	$-2\sqrt{3} \approx -3.4$

θ	$r=4\sin\theta$	θ	$r=4\sin\theta$
$\frac{\pi}{2}$	4	$\frac{3\pi}{2}$	-4
$\frac{2\pi}{3}$	$2\sqrt{3}\approx 3.4$	$\frac{5\pi}{3}$	$-2\sqrt{3}\approx -3.4$
$\frac{3\pi}{4}$	$2\sqrt{2}\approx 2.8$	$\frac{7\pi}{4}$	$-2\sqrt{2}\approx -2.8$
$\frac{5\pi}{6}$	2	$\frac{11\pi}{6}$	-2
		2π	0

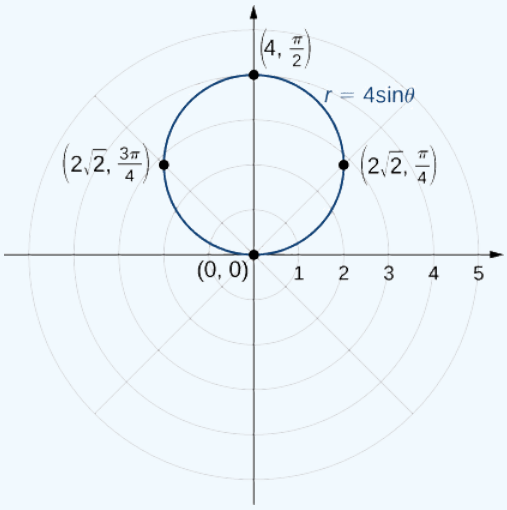


Figure 7.3.4: The graph of the function $r=4\sin\theta$ is a circle.

This is the graph of a circle. The equation $r=4\sin\theta$ can be converted into rectangular coordinates by first multiplying both sides by r. This gives the equation $r^2=4r\sin\theta$. Next use the facts that $r^2=x^2+y^2$ and $y=r\sin\theta$. This gives $x^2+y^2=4y$. To put this equation into standard form, subtract $4y$ from both sides of the equation and complete the square:

$$\begin{aligned} x^2+y^2-4y &= 0 \\ x^2+(y^2-4y) &= 0 \\ x^2+(y^2-4y+4) &= 0+4 \\ x^2+(y-2)^2 &= 4 \end{aligned}$$

This is the equation of a circle with radius 2 and center $(0,2)$ in the rectangular coordinate system.

? Exercise 7.3.3

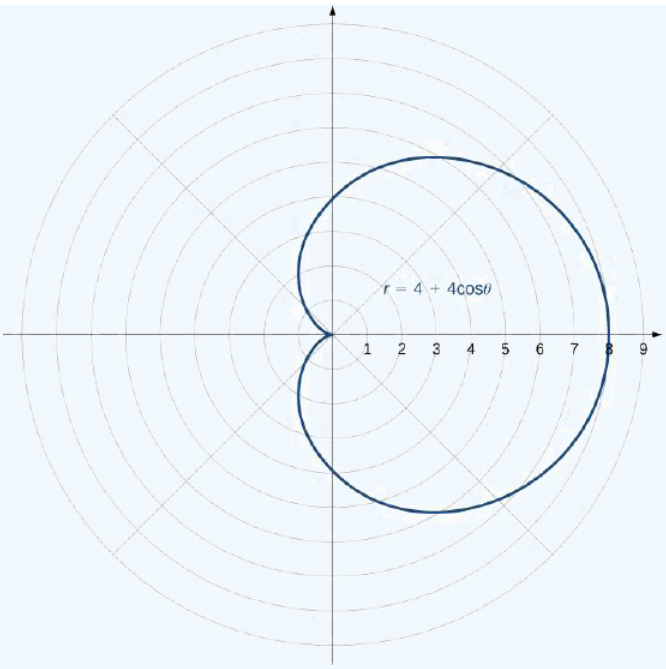
Create a graph of the curve defined by the function $r=4+4\cos\theta$.

Hint

Follow the problem-solving strategy for creating a graph in polar coordinates.

Answer

The name of this shape is a cardioid, which we will study further later in this section.



The graph in Example 7.3.3 was that of a circle. The equation of the circle can be transformed into rectangular coordinates using the coordinate transformation formulas in Equation 7.3.1. Example 7.3.4 gives some more examples of functions for transforming from polar to rectangular coordinates.

Example 7.3.4: Transforming Polar Equations to Rectangular Coordinates

Rewrite each of the following equations in rectangular coordinates and identify the graph.

a. $\theta = \dfrac{\pi}{3}$
b. $r = 3$
c. $r = 6\cos\theta - 8\sin\theta$

Solution:

a. Take the tangent of both sides. This gives $\tan\theta = \tan(\pi/3) = \sqrt{3}$. Since $\tan\theta = y/x$ we can replace the left-hand side of this equation by y/x. This gives $y/x = \sqrt{3}$, which can be rewritten as $y = x\sqrt{3}$. This is the equation of a straight line passing through the origin with slope $\sqrt{3}$. In general, any polar equation of the form $\theta = K$ represents a straight line through the pole with slope equal to $\tan K$.

b. First, square both sides of the equation. This gives $r^2 = 9$. Next replace r^2 with $x^2 + y^2$. This gives the equation $x^2 + y^2 = 9$, which is the equation of a circle centered at the origin with radius 3. In general, any polar equation of the form $r = k$ where *k* is a positive constant represents a circle of radius *k* centered at the origin. (Note: when squaring both sides of an equation it is possible to introduce new points unintentionally. This should always be taken into consideration. However, in this case we do not introduce new points. For example, $(-3, \dfrac{\pi}{3})$ is the same point as $(3, \dfrac{4\pi}{3})$.)

c. Multiply both sides of the equation by r. This leads to $r^2 = 6r\cos\theta - 8r\sin\theta$. Next use the formulas

$$r^2 = x^2 + y^2,\ x = r\cos\theta,\ y = r\sin\theta.$$

This gives

$$r^2 = 6(r\cos\theta) - 8(r\sin\theta)$$
$$x^2 + y^2 = 6x - 8y.$$

To put this equation into standard form, first move the variables from the right-hand side of the equation to the left-hand side, then complete the square.

$$x^2+y^2=6x-8y$$
$$x^2-6x+y^2+8y=0$$
$$(x^2-6x)+(y^2+8y)=0$$
$$(x^2-6x+9)+(y^2+8y+16)=9+16$$
$$(x-3)^2+(y+4)^2=25.$$

This is the equation of a circle with center at $(3,-4)$ and radius 5. Notice that the circle passes through the origin since the center is 5 units away.

Exercise 7.3.4

Rewrite the equation $r=\sec\theta\tan\theta$ in rectangular coordinates and identify its graph.

Hint

Convert to sine and cosine, then multiply both sides by cosine.

Answer

$y=x^2$, which is the equation of a parabola opening upward.

We have now seen several examples of drawing graphs of curves defined by **polar equations.** A summary of some common curves is given in the tables below. In each equation, *a* and *b* are arbitrary constants.

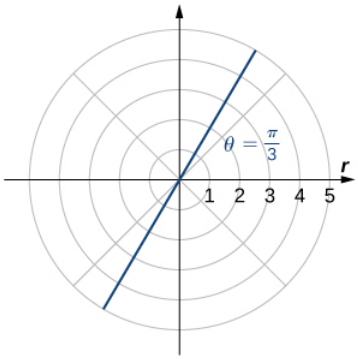
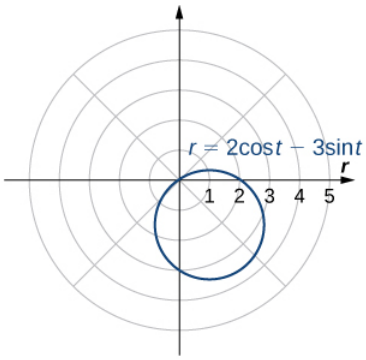
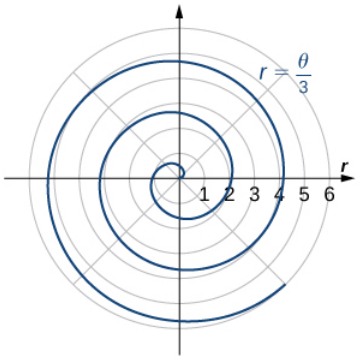
Name	Equation	Example
Line passing through the pole with slope tan K	$\theta = K$	
Circle	$r = a\cos\theta + b\sin\theta$	
Spiral	$r = a + b\theta$	

Figure 7.3.5

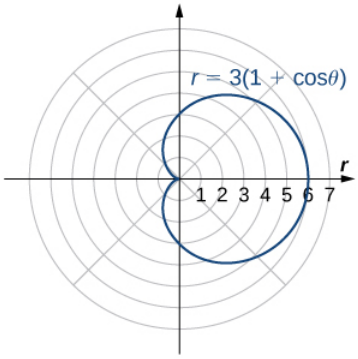
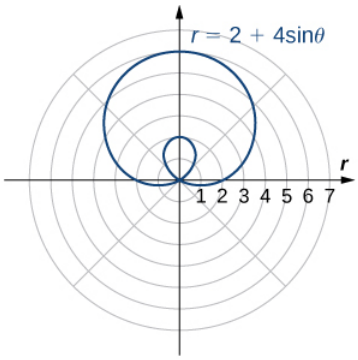
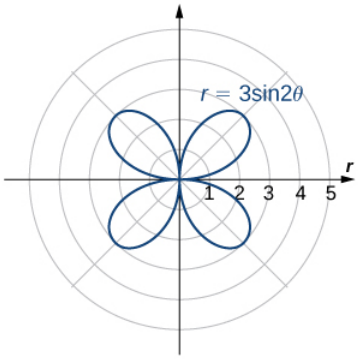
Name	Equation	Example
Cardioid	$r = a(1 + \cos\theta)$ $r = a(1 - \cos\theta)$ $r = a(1 + \sin\theta)$ $r = a(1 - \sin\theta)$	
Limaçon	$r = a\cos\theta + b$ $r = a\sin\theta + b$	
Rose	$r = a\cos(b\theta)$ $r = a\sin(b\theta)$	

Figure 7.3.6

A **cardioid** is a special case of a **limaçon** (pronounced “lee-mah-son”), in which $a = b$ or $a = -b$. The **rose i**s a very interesting curve. Notice that the graph of $r = 3\sin 2\theta$ has four petals. However, the graph of $r = 3\sin 3\theta$ has three petals as shown.

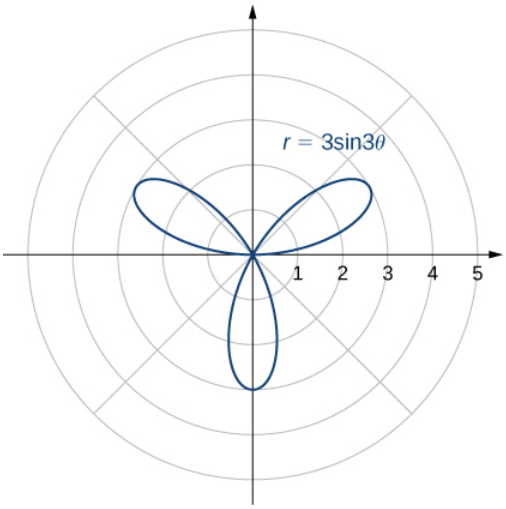


Figure 7.3.7: Graph of $r = 3\sin 3\theta$.

If the coefficient of θ is even, the graph has twice as many petals as the coefficient. If the coefficient of θ is odd, then the number of petals equals the coefficient. You are encouraged to explore why this happens. Even more interesting graphs emerge when the coefficient of θ is not an integer. For example, if it is rational, then the curve is closed; that is, it eventually ends where it started (Figure 7.3.8*a*). However, if the coefficient is irrational, then the curve never closes (Figure 7.3.8*b*). Although it may appear that the curve is closed, a closer examination reveals that the petals just above the positive x axis are slightly thicker. This is because the petal does not quite match up with the starting point.

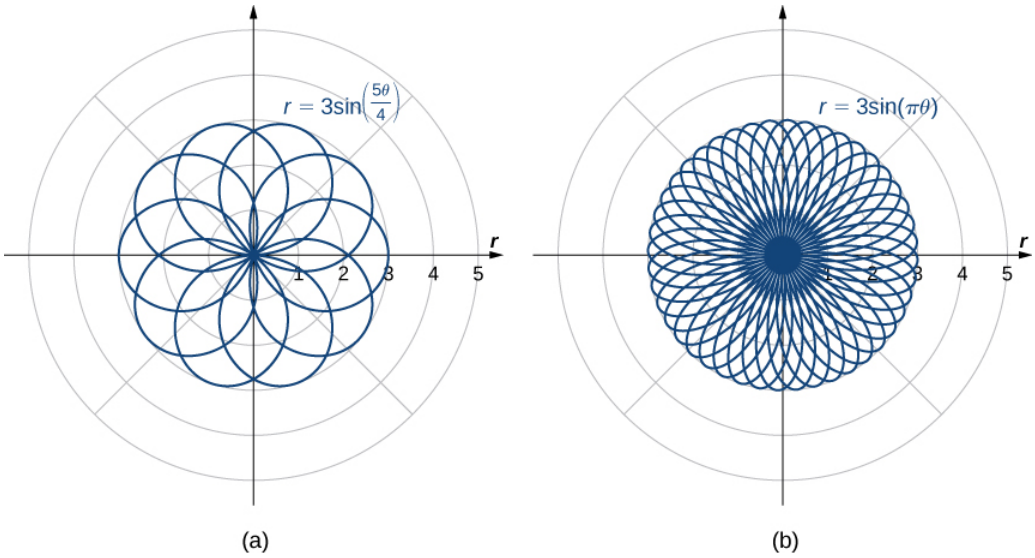


Figure 7.3.8: Polar rose graphs of functions with (a) rational coefficient and (b) irrational coefficient. Note that the rose in part (b) would actually fill the entire circle if plotted in full.

Since the curve defined by the graph of $r = 3\sin(\pi\theta)$ never closes, the curve depicted in Figure 7.3.8*b* is only a partial depiction. In fact, this is an example of a **space-filling curve.** A space-filling curve is one that in fact occupies a two-dimensional subset of the real plane. In this case the curve occupies the circle of radius 3 centered at the origin.

Example 7.3.5: Describing a Spiral

Recall the **chambered nautilus** introduced in the chapter prelude. This creature displays a spiral when half the outer shell is cut away. It is possible to describe a spiral using rectangular coordinates. Figure 7.3.9 shows a spiral in rectangular coordinates. How can we describe this curve mathematically?

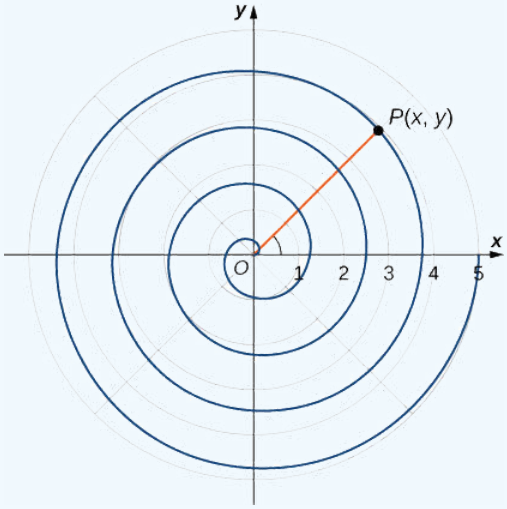


Figure 7.3.9: How can we describe a spiral graph mathematically?

Solution

As the point P travels around the spiral in a counterclockwise direction, its distance d from the origin increases. Assume that the distance d is a constant multiple k of the angle θ that the line segment OP makes with the positive x-axis. Therefore $d(P,O)=k\theta$, where O is the origin. Now use the distance formula and some trigonometry:

$$d(P,O)=k\theta$$

$$\sqrt{(x-0)^2+(y-0)^2}=k\arctan(\frac{y}{x})$$

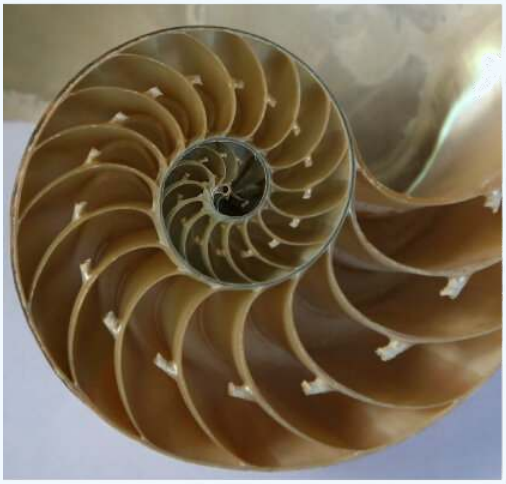
$$\sqrt{x^2+y^2}=k\arctan(\frac{y}{x})$$

$$\arctan(\frac{y}{x})=\frac{\sqrt{x^2+y^2}}{k}$$

$$y=x\tan(\frac{\sqrt{x^2+y^2}}{k}).$$

Although this equation describes the spiral, it is not possible to solve it directly for either x or y. However, if we use polar coordinates, the equation becomes much simpler. In particular, $d(P,O)=r$, and θ is the second coordinate. Therefore the equation for the spiral becomes $r=k\theta$. Note that when $\theta=0$ we also have $r=0$, so the spiral emanates from the origin. We can remove this restriction by adding a constant to the equation. Then the equation for the spiral becomes $r=a+k\theta$ for arbitrary constants a and k. This is referred to as an **Archimedean spiral**, after the Greek mathematician Archimedes.

Another type of spiral is the logarithmic spiral, described by the function $r=a\cdot b^{\theta}$. A graph of the function $r=1.2(1.25^{\theta})$ is given in Figure 7.3.10. This spiral describes the shell shape of the chambered nautilus.



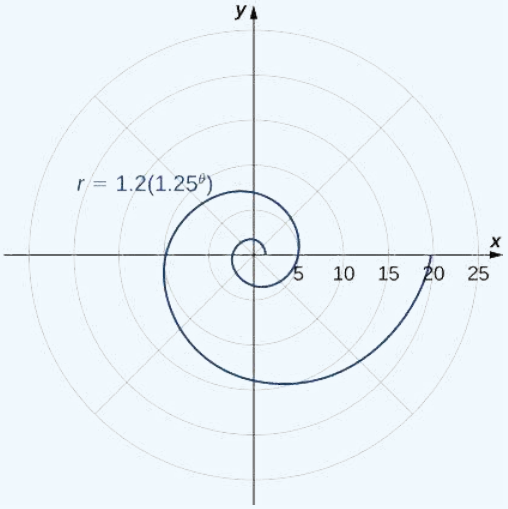


Figure 7.3.10: A logarithmic spiral is similar to the shape of the chambered nautilus shell. (credit: modification of work by Jitze Couperus, Flickr)

Suppose a curve is described in the polar coordinate system via the function $r = f(\theta)$. Since we have conversion formulas from polar to rectangular coordinates given by

$$x = r\cos\theta$$

$$y = r\sin\theta$$

,

it is possible to rewrite these formulas using the function

$$x = f(\theta)\cos\theta$$

$$y = f(\theta)\sin\theta.$$

This step gives a parameterization of the curve in rectangular coordinates using θ as the parameter. For example, the spiral formula $r = a + b\theta$ from Figure becomes

$$x = (a + b\theta)\cos\theta$$

$$y = (a + b\theta)\sin\theta.$$

Letting θ range from $-\infty$ to ∞ generates the entire spiral.

Symmetry in Polar Coordinates

When studying **symmetry** of functions in rectangular coordinates (i.e., in the form $y = f(x)$), we talk about symmetry with respect to the *y*-axis and symmetry with respect to the origin. In particular, if $f(-x) = f(x)$ for all x in the domain of f, then f is an even function and its graph is symmetric with respect to the *y*-axis. If $f(-x) = -f(x)$ for all *x* in the domain of f, then f is an odd function and its graph is symmetric with respect to the origin. By determining which types of symmetry a graph exhibits, we can learn more about the shape and appearance of the graph. Symmetry can also reveal other properties of the function that generates the graph. Symmetry in polar curves works in a similar fashion.

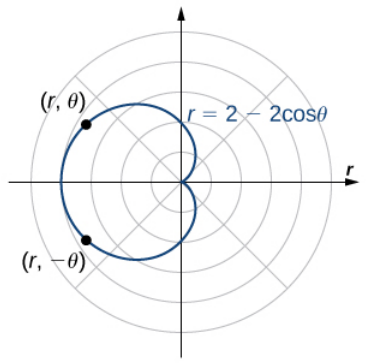
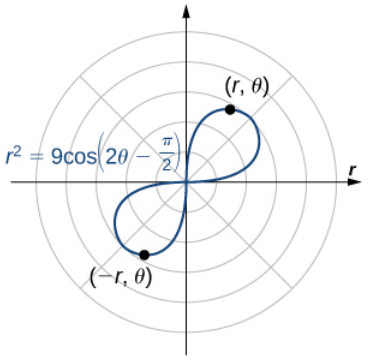
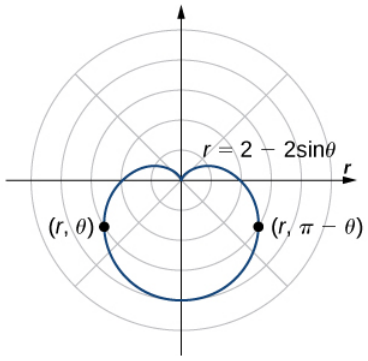
Symmetry in Polar Curves and Equations

Consider a curve generated by the function $r = f(\theta)$ in polar coordinates.

i. The curve is symmetric about the polar axis if for every point (r, θ) on the graph, the point $(r, -\theta)$ is also on the graph. Similarly, the equation $r = f(\theta)$ is unchanged by replacing θ with $-\theta$.
ii. The curve is symmetric about the pole if for every point (r, θ) on the graph, the point $(r, \pi + \theta)$ is also on the graph. Similarly, the equation $r = f(\theta)$ is unchanged when replacing r with $-r$, or θ with $\pi + \theta$.

iii. The curve is symmetric about the vertical line $\theta = \frac{\pi}{2}$ if for every point (r, θ) on the graph, the point $(r, \pi - \theta)$ is also on the graph. Similarly, the equation $r = f(\theta)$ is unchanged when θ is replaced by $\pi - \theta$.

The following table shows examples of each type of symmetry.

Symmetry with respect to the polar axis: For every point (r, θ) on the graph, there is also a point reflected directly across the horizontal (polar) axis.	
Symmetry with respect to the pole: For every point (r, θ) on the graph, there is also a point on the graph that is reflected through the pole as well.	
Symmetry with respect to the vertical line $\theta = \frac{\pi}{2}$: For every point (r, θ) on the graph, there is also a point reflected directly across the vertical axis.	

Example 7.3.6: Using Symmetry to Graph a Polar Equation

Find the symmetry of the rose defined by the equation $r = 3\sin(2\theta)$ and create a graph.

Solution

Suppose the point (r, θ) is on the graph of $r = 3\sin(2\theta)$.

i. To test for symmetry about the polar axis, first try replacing θ with $-\theta$. This gives $r = 3\sin(2(-\theta)) = -3\sin(2\theta)$. Since this changes the original equation, this test is not satisfied. However, returning to the original equation and replacing r with $-r$ and θ with $\pi - \theta$ yields

$$\begin{aligned} -r &= 3\sin(2(\pi-\theta)) \\ -r &= 3\sin(2\pi-2\theta) \\ -r &= 3\sin(-2\theta) \\ -r &= -3\sin 2\theta. \end{aligned}$$

Multiplying both sides of this equation by -1 gives $r = 3\sin 2\theta$, which is the original equation. This demonstrates that the graph is symmetric with respect to the polar axis.

ii. To test for symmetry with respect to the pole, first replace r with $-r$, which yields $-r = 3\sin(2\theta)$. Multiplying both sides by -1 gives $r = -3\sin(2\theta)$, which does not agree with the original equation. Therefore the equation does not pass the test for this symmetry. However, returning to the original equation and replacing θ with $\theta+\pi$ gives

$$\begin{aligned} r &= 3\sin(2(\theta+\pi)) \\ &= 3\sin(2\theta+2\pi) \\ &= 3(\sin 2\theta\cos 2\pi+\cos 2\theta\sin 2\pi) \\ &= 3\sin 2\theta. \end{aligned}$$

Since this agrees with the original equation, the graph is symmetric about the pole.

iii. To test for symmetry with respect to the vertical line $\theta = \dfrac{\pi}{2}$, first replace both r with $-r$ and θ with $-\theta$.

$$\begin{aligned} -r &= 3\sin(2(-\theta)) \\ -r &= 3\sin(-2\theta) \\ -r &= -3\sin 2\theta. \end{aligned}$$

Multiplying both sides of this equation by -1 gives $r = 3\sin 2\theta$, which is the original equation. Therefore the graph is symmetric about the vertical line $\theta = \dfrac{\pi}{2}$.

This graph has symmetry with respect to the polar axis, the origin, and the vertical line going through the pole. To graph the function, tabulate values of θ between 0 and $\pi/2$ and then reflect the resulting graph.

Table 7.3.2: Table of values for this polar function.

θ	$r = 3\sin(2\theta)$
0	0
$\dfrac{\pi}{6}$	$\dfrac{3\sqrt{3}}{2} \approx 2.6$
$\dfrac{\pi}{4}$	3
$\dfrac{\pi}{3}$	$\dfrac{3\sqrt{3}}{2} \approx 2.6$
$\dfrac{\pi}{2}$	0

This gives one petal of the rose, as shown in the following graph.

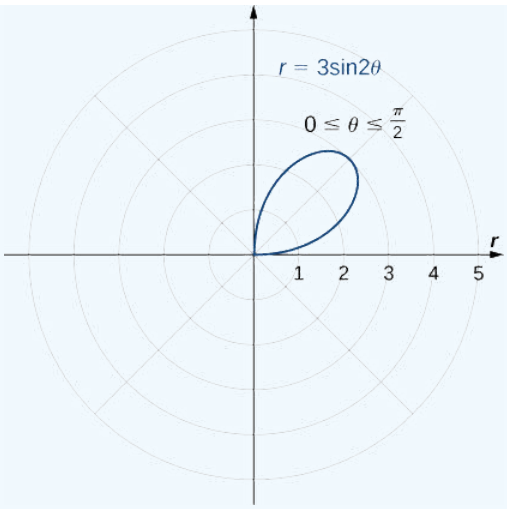


Figure 7.3.11: The graph of the equation between $\theta = 0$ and $\theta = \pi/2$.

Reflecting this image into the other three quadrants gives the entire graph as shown.

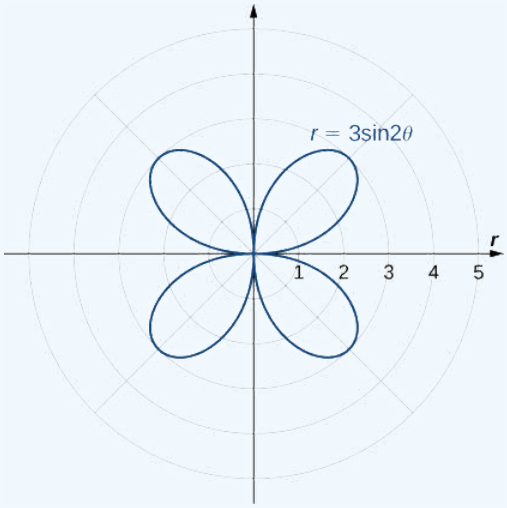


Figure 7.3.12: The entire graph of the equation is called a four-petaled rose.

? Exercise 7.3.5Symmetry

Determine the symmetry of the graph determined by the equation $r = 2\cos(3\theta)$ and create a graph.

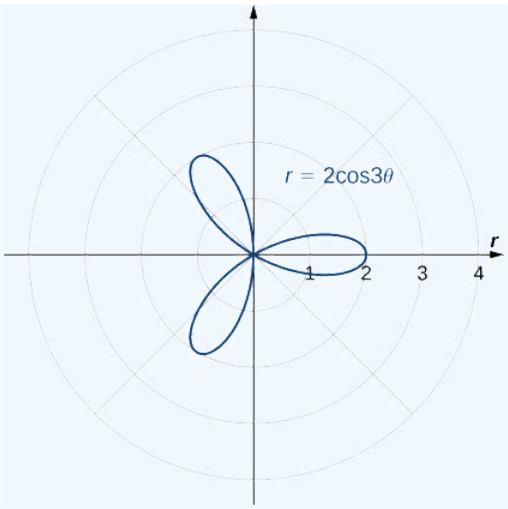
Hint

Use Note.

Answer

Symmetric with respect to the polar axis.

https://math.libretexts.org/@go/page/33835



Key Concepts

- The polar coordinate system provides an alternative way to locate points in the plane.
- Convert points between rectangular and polar coordinates using the formulas

$$x = r\cos\theta \text{ and } y = r\sin\theta$$

and

$$r = \sqrt{x^2 + y^2} \text{ and } \tan\theta = \frac{y}{x}.$$

- To sketch a polar curve from a given polar function, make a table of values and take advantage of periodic properties.
- Use the conversion formulas to convert equations between rectangular and polar coordinates.
- Identify symmetry in polar curves, which can occur through the pole, the horizontal axis, or the vertical axis.

Glossary

angular coordinate

θ the angle formed by a line segment connecting the origin to a point in the polar coordinate system with the positive radial x-axis, measured counterclockwise

cardioid

a plane curve traced by a point on the perimeter of a circle that is rolling around a fixed circle of the same radius; the equation of a cardioid is $r = a(1 + \sin\theta)$ or $r = a(1 + \cos\theta)$

limaçon

the graph of the equation $r = a + b\sin\theta$ or $r = a + b\cos\theta$. If $a = b$ then the graph is a cardioid

polar axis

the horizontal axis in the polar coordinate system corresponding to $r \geq 0$

polar coordinate system

a system for locating points in the plane. The coordinates are r, the radial coordinate, and θ, the angular coordinate

polar equation

an equation or function relating the radial coordinate to the angular coordinate in the polar coordinate system

pole

the central point of the polar coordinate system, equivalent to the origin of a Cartesian system

radial coordinate

r the coordinate in the polar coordinate system that measures the distance from a point in the plane to the pole

rose

graph of the polar equation $r = a\cos 2\theta$ or $r = a\sin 2\theta$ for a positive constant a

space-filling curve

a curve that completely occupies a two-dimensional subset of the real plane

---

This page titled 7.3: Polar Coordinates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **11.3: Polar Coordinates** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

7.4: Area and Arc Length in Polar Coordinates

Learning Objectives

- Apply the formula for area of a region in polar coordinates.
- Determine the arc length of a polar curve.

In the rectangular coordinate system, the definite integral provides a way to calculate the area under a curve. In particular, if we have a function $y = f(x)$ defined from $x = a$ to $x = b$ where $f(x) > 0$ on this interval, the area between the curve and the x-axis is given by

$$A = \int_a^b f(x)dx.$$

This fact, along with the formula for evaluating this integral, is summarized in the Fundamental Theorem of Calculus. Similarly, the arc length of this curve is given by

$$L = \int_a^b \sqrt{1 + (f'(x))^2}dx.$$

In this section, we study analogous formulas for area and arc length in the polar coordinate system.

Areas of Regions Bounded by Polar Curves

We have studied the formulas for area under a curve defined in rectangular coordinates and parametrically defined curves. Now we turn our attention to deriving a formula for the area of a region bounded by a polar curve. Recall that the proof of the Fundamental Theorem of Calculus used the concept of a Riemann sum to approximate the area under a curve by using rectangles. For polar curves we use the Riemann sum again, but the rectangles are replaced by sectors of a circle.

Consider a curve defined by the function $r = f(\theta)$, where $\alpha \leq \theta \leq \beta$. Our first step is to partition the interval $[\alpha, \beta]$ into n equal-width subintervals. The width of each subinterval is given by the formula $\Delta\theta = (\beta - \alpha)/n$, and the ith partition point θ_i is given by the formula $\theta_i = \alpha + i\Delta\theta$. Each partition point $\theta = \theta_i$ defines a line with slope $\tan\theta_i$ passing through the pole as shown in the following graph.

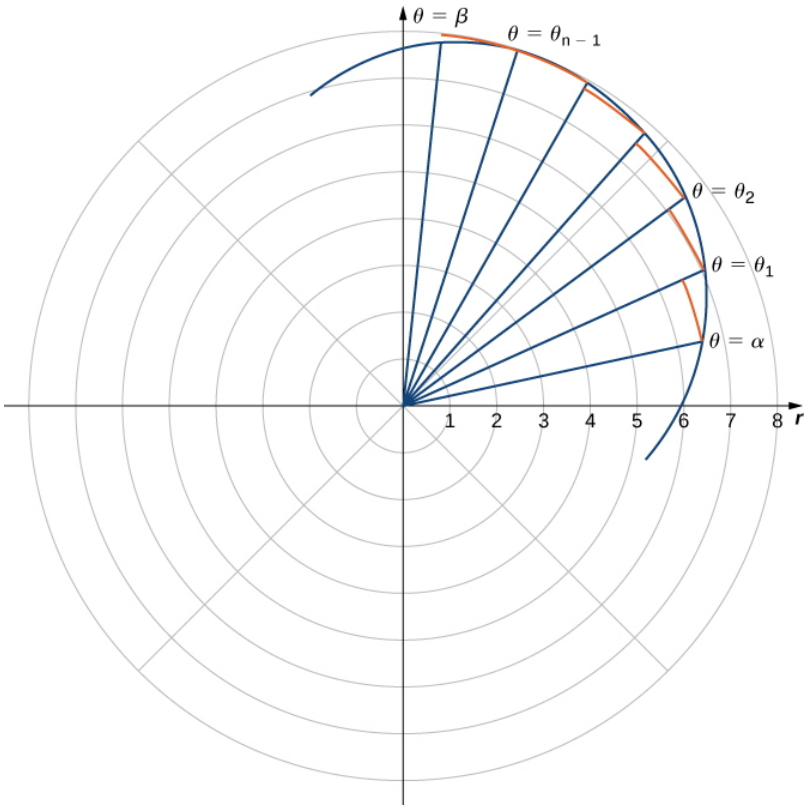


Figure 7.4.1: A partition of a typical curve in polar coordinates.

The line segments are connected by arcs of constant radius. This defines sectors whose areas can be calculated by using a geometric formula. The area of each sector is then used to approximate the area between successive line segments. We then sum the areas of the sectors to approximate the total area. This approach gives a Riemann sum approximation for the total area. The formula for the area of a sector of a circle is illustrated in the following figure.

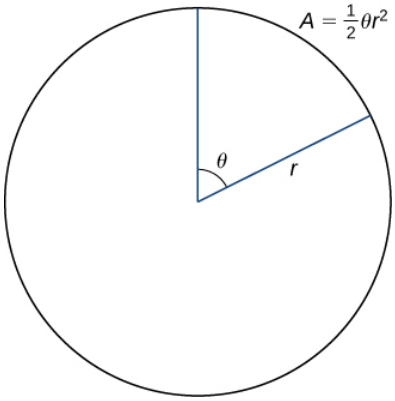


Figure 7.4.2: *The area of a sector of a circle is given by* $A = \frac{1}{2}\theta r^2$.

Recall that the area of a circle is $A = \pi r^2$. When measuring angles in radians, 360 degrees is equal to 2π radians. Therefore a fraction of a circle can be measured by the central angle θ. The fraction of the circle is given by $\frac{\theta}{2\pi}$, so the area of the sector is this fraction multiplied by the total area:

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

Since the radius of a typical sector in Figure 7.4.1 is given by $r_i = f(\theta_i)$, the area of the ith sector is given by

$$A_i = \frac{1}{2}(\Delta\theta)(f(\theta_i))^2.$$

Therefore a Riemann sum that approximates the area is given by

$$A_n = \sum_{i=1}^{n} A_i \approx \sum_{i=1}^{n} \frac{1}{2}(\Delta\theta)(f(\theta_i))^2.$$

We take the limit as $n \to \infty$ to get the exact area:

$$A = \lim_{n\to\infty} A_n = \frac{1}{2}\int_{\alpha}^{\beta} (f(\theta))^2 d\theta.$$

This gives the following theorem.

Area of a Region Bounded by a Polar Curve

Suppose f is continuous and nonnegative on the interval $\alpha \le \theta \le \beta$ with $0 < \beta - \alpha \le 2\pi$. The area of the region bounded by the graph of $r = f(\theta)$ between the radial lines $\theta = \alpha$ and $\theta = \beta$ is

$$A = \frac{1}{2}\int_{\alpha}^{\beta} [f(\theta)]^2 d\theta \tag{7.4.1}$$

$$= \frac{1}{2}\int_{\alpha}^{\beta} r^2 d\theta. \tag{7.4.2}$$

Example 7.4.1: Finding an Area of a Polar Region

Find the area of one petal of the rose defined by the equation $r = 3\sin(2\theta)$.

Solution

The graph of $r = 3\sin(2\theta)$ follows.

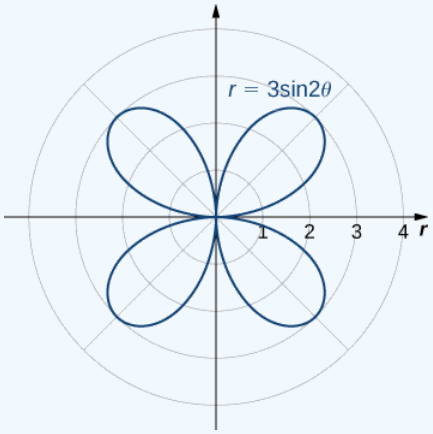


Figure 7.4.3: The graph of $r = 3\sin(2\theta)$.

When $\theta = 0$ we have $r = 3\sin(2(0)) = 0$. The next value for which $r = 0$ is $\theta = \pi/2$. This can be seen by solving the equation $3\sin(2\theta) = 0$ for θ. Therefore the values $\theta = 0$ to $\theta = \pi/2$ trace out the first petal of the rose. To find the area inside this petal, use Equation 7.4.2 with $f(\theta) = 3\sin(2\theta)$, $\alpha = 0$, and $\beta = \pi/2$:

$$A = \frac{1}{2}\int_{\alpha}^{\beta} [f(\theta)]^2 d\theta$$

$$= \frac{1}{2}\int_{0}^{\pi/2} [3\sin(2\theta)]^2 d\theta$$

$$= \frac{1}{2}\int_{0}^{\pi/2} 9\sin^2(2\theta) d\theta.$$

To evaluate this integral, use the formula $\sin^2\alpha = (1 - \cos(2\alpha))/2$ with $\alpha = 2\theta$:

$$
\begin{aligned}
A &= \frac{1}{2}\int_0^{\pi/2} 9\sin^2(2\theta)d\theta \\
&= \frac{9}{2}\int_0^{\pi/2} \frac{(1-\cos(4\theta))}{2}d\theta \\
&= \frac{9}{4}(\int_0^{\pi/2} 1-\cos(4\theta)d\theta) \\
&= \frac{9}{4}(\theta - \frac{\sin(4\theta)}{4}\Big|_0^{\pi/2} \\
&= \frac{9}{4}(\frac{\pi}{2} - \frac{\sin 2\pi}{4}) - \frac{9}{4}(0 - \frac{\sin 4(0)}{4}) \\
&= \frac{9\pi}{8}
\end{aligned}
$$

? Exercise 7.4.1

Find the area inside the cardioid defined by the equation $r = 1 - \cos\theta$.

Hint

Use Equation 7.4.2 Be sure to determine the correct limits of integration before evaluating.

Answer

$A = 3\pi/2$

Example 7.4.1 involved finding the area inside one curve. We can also use Equation 7.4.2 to find the area between two polar curves. However, we often need to find the points of intersection of the curves and determine which function defines the outer curve or the inner curve between these two points.

✓ Example 7.4.2: Finding the Area between Two Polar Curves

Find the area outside the cardioid $r = 2 + 2\sin\theta$ and inside the circle $r = 6\sin\theta$.

Solution

First draw a graph containing both curves as shown.

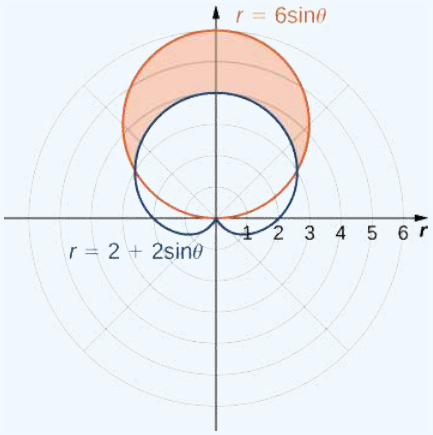


Figure 7.4.4: *The region between the curves* $r = 2 + 2\sin\theta$ *and* $r = 6\sin\theta$.

To determine the limits of integration, first find the points of intersection by setting the two functions equal to each other and solving for θ:

$$6\sin\theta = 2 + 2\sin\theta$$
$$4\sin\theta = 2$$
$$\sin\theta = \frac{1}{2}.$$

This gives the solutions $\theta = \frac{\pi}{6}$ and $\theta = \frac{5\pi}{6}$, which are the limits of integration. The circle $r = 3\sin\theta$ is the red graph, which is the outer function, and the cardioid $r = 2 + 2\sin\theta$ is the blue graph, which is the inner function. To calculate the area between the curves, start with the area inside the circle between $\theta = \frac{\pi}{6}$ and $\theta = \frac{5\pi}{6}$, then subtract the area inside the cardioid between $\theta = \frac{\pi}{6}$ and $\theta = \frac{5\pi}{6}$:

$$\begin{aligned} A &= \text{circle} - \text{cardioid} \\ &= \frac{1}{2}\int_{\pi/6}^{5\pi/6}[6\sin\theta]^2 d\theta - \frac{1}{2}\int_{\pi/6}^{5\pi/6}[2+2\sin\theta]^2 d\theta \\ &= \frac{1}{2}\int_{\pi/6}^{5\pi/6} 36\sin^2\theta\, d\theta - \frac{1}{2}\int_{\pi/6}^{5\pi/6} 4 + 8\sin\theta + 4\sin^2\theta\, d\theta \\ &= 18\int_{\pi/6}^{5\pi/6} \frac{1-\cos(2\theta)}{2} d\theta - 2\int_{\pi/6}^{5\pi/6} 1 + 2\sin\theta + \frac{1-\cos(2\theta)}{2} d\theta \\ &= 9[\theta - \frac{\sin(2\theta)}{2}]_{\pi/6}^{5\pi/6} - 2[\frac{3\theta}{2} - 2\cos\theta - \frac{\sin(2\theta)}{4}]_{\pi/6}^{5\pi/6} \\ &= 9(\frac{5\pi}{6} - \frac{\sin(10\pi/6)}{2}) - 9(\frac{\pi}{6} - \frac{\sin(2\pi/6)}{2}) - (3(\frac{5\pi}{6}) - 4\cos\frac{5\pi}{6} - \frac{\sin(10\pi/6)}{2}) \\ &+ (3(\frac{\pi}{6}) - 4\cos\frac{\pi}{6} - \frac{\sin(2\pi/6)}{2}) \\ &= 4\pi. \end{aligned}$$

Exercise 7.4.2

Find the area inside the circle $r = 4\cos\theta$ and outside the circle $r = 2$.

Hint

Use Equation 7.4.2 and take advantage of symmetry.

Answer

$$A = \frac{4\pi}{3} + 2\sqrt{3}$$

In Example 7.4.2 we found the area inside the circle and outside the cardioid by first finding their intersection points. Notice that solving the equation directly for θ yielded two solutions: $\theta = \frac{\pi}{6}$ and $\theta = \frac{5\pi}{6}$. However, in the graph there are three intersection points. The third intersection point is the origin. The reason why this point did not show up as a solution is because the origin is on both graphs but for different values of θ. For example, for the cardioid we get

$$2 + 2\sin\theta = 0$$
$$\sin\theta = -1,$$

so the values for θ that solve this equation are $\theta = \frac{3\pi}{2} + 2n\pi$, where n is any integer. For the circle we get

$$6\sin\theta = 0.$$

The solutions to this equation are of the form $\theta = n\pi$ for any integer value of n. These two solution sets have no points in common. Regardless of this fact, the curves intersect at the origin. This case must always be taken into consideration.

Arc Length in Polar Curves

Here we derive a formula for the arc length of a curve defined in polar coordinates. In rectangular coordinates, the arc length of a parameterized curve $(x(t), y(t))$ for $a \le t \le b$ is given by

$$L = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

In polar coordinates we define the curve by the equation $r = f(\theta)$, where $\alpha \le \theta \le \beta$. In order to adapt the arc length formula for a polar curve, we use the equations

$$x = r\cos\theta = f(\theta)\cos\theta$$

and

$$y = r\sin\theta = f(\theta)\sin\theta,$$

and we replace the parameter t by θ. Then

$$\frac{dx}{d\theta} = f'(\theta)\cos\theta - f(\theta)\sin\theta$$

$$\frac{dy}{d\theta} = f'(\theta)\sin\theta + f(\theta)\cos\theta.$$

We replace dt by $d\theta$, and the lower and upper limits of integration are α and β, respectively. Then the arc length formula becomes

$$\begin{aligned} L &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \\ &= \int_\alpha^\beta \sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2}\, d\theta \\ &= \int_\alpha^\beta \sqrt{(f'(\theta)\cos\theta - f(\theta)\sin\theta)^2 + (f'(\theta)\sin\theta + f(\theta)\cos\theta)^2}\, d\theta \\ &= \int_\alpha^\beta \sqrt{(f'(\theta))^2(\cos^2\theta + \sin^2\theta) + (f(\theta))^2(\cos^2\theta + \sin^2\theta)}\, d\theta \\ &= \int_\alpha^\beta \sqrt{(f'(\theta))^2 + (f(\theta))^2}\, d\theta \\ &= \int_\alpha^\beta \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta \end{aligned}$$

This gives us the following theorem.

Arc Length of a Curve Defined by a Polar Function

Let f be a function whose derivative is continuous on an interval $\alpha \le \theta \le \beta$. The arc length of the graph of $r = f(\theta)$ from $\theta = \alpha$ to $\theta = \beta$ is

$$L = \int_\alpha^\beta \sqrt{[f(\theta)]^2 + [f'(\theta)]^2}\, d\theta \tag{7.4.3}$$

$$= \int_\alpha^\beta \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta. \tag{7.4.4}$$

Example 7.4.3: Finding the Arc Length of a cardioid

Find the arc length of the cardioid $r = 2 + 2\cos\theta$.

Solution

When $\theta = 0, r = 2 + 2\cos 0 = 4$. Furthermore, as θ goes from 0 to 2π, the cardioid is traced out exactly once. Therefore these are the limits of integration. Using $f(\theta) = 2 + 2\cos\theta, \alpha = 0,$ and $\beta = 2\pi$, Equation 7.4.3 becomes

$$\begin{aligned} L &= \int_{\alpha}^{\beta} \sqrt{[f(\theta)]^2 + [f'(\theta)]^2}\, d\theta \\ &= \int_0^{2\pi} \sqrt{[2+2\cos\theta]^2 + [-2\sin\theta]^2}\, d\theta \\ &= \int_0^{2\pi} \sqrt{4 + 8\cos\theta + 4\cos^2\theta + 4\sin^2\theta}\, d\theta \\ &= \int_0^{2\pi} \sqrt{4 + 8\cos\theta + 4(\cos^2\theta + \sin^2\theta)}\, d\theta \\ &= \int_0^{2\pi} \sqrt{8 + 8\cos\theta}\, d\theta \\ &= 2\int_0^{2\pi} \sqrt{2 + 2\cos\theta}\, d\theta. \end{aligned}$$

Next, using the identity $\cos(2\alpha) = 2\cos^2\alpha - 1,$ add 1 to both sides and multiply by 2. This gives $2 + 2\cos(2\alpha) = 4\cos^2\alpha$. Substituting $\alpha = \theta/2$ gives $2 + 2\cos\theta = 4\cos^2(\theta/2)$, so the integral becomes

$$\begin{aligned} L &= 2\int_0^{2\pi} \sqrt{2 + 2\cos\theta}\, d\theta \\ &= 2\int_0^{2\pi} \sqrt{4\cos^2(\frac{\theta}{2})}\, d\theta \\ &= 4\int_0^{2\pi} |\cos(\frac{\theta}{2})|\, d\theta. \end{aligned}$$

The absolute value is necessary because the cosine is negative for some values in its domain. To resolve this issue, change the limits from 0 to π and double the answer. This strategy works because cosine is positive between 0 and $\frac{\pi}{2}$. Thus,

$$\begin{aligned} L &= 4\int_0^{2\pi} |\cos(\frac{\theta}{2})|\, d\theta \\ &= 8\int_0^{\pi} \cos(\frac{\theta}{2})\, d\theta \\ &= 8(2\sin(\frac{\theta}{2})\Big|_0^{\pi} \\ &= 16 \end{aligned}$$

Exercise 7.4.3

Find the total arc length of the graph of $r = 3\sin\theta$.

Hint

Use Equation 7.4.3 To determine the correct limits, make a table of values.

Answer

$s = 3\pi$

Key Concepts

- The area of a region in polar coordinates defined by the equation $r = f(\theta)$ with $\alpha \leq \theta \leq \beta$ is given by the integral $A = \frac{1}{2}\int_{\alpha}^{\beta}[f(\theta)]^2 d\theta$.
- To find the area between two curves in the polar coordinate system, first find the points of intersection, then subtract the corresponding areas.
- The arc length of a polar curve defined by the equation $r = f(\theta)$ with $\alpha \leq \theta \leq \beta$ is given by the integral $L = \int_{\alpha}^{\beta}\sqrt{[f(\theta)]^2 + [f\,'(\theta)]^2}\,d\theta = \int_{\alpha}^{\beta}\sqrt{r^2 + (\frac{dr}{d\theta})^2}\,d\theta$.

Key Equations

- **Area of a region bounded by a polar curve**

$$A = \frac{1}{2}\int_{\alpha}^{\beta}[f(\theta)]^2 d\theta = \frac{1}{2}\int_{\alpha}^{\beta} r^2 d\theta$$

- **Arc length of a polar curve**

$$L = \int_{\alpha}^{\beta}\sqrt{[f(\theta)]^2 + [f\,'(\theta)]^2}\,d\theta = \int_{\alpha}^{\beta}\sqrt{r^2 + (\frac{dr}{d\theta})^2}\,d\theta$$

This page titled 7.4: Area and Arc Length in Polar Coordinates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **11.4: Area and Arc Length in Polar Coordinates** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

7.5: Conic Sections

Learning Objectives

- Identify the equation of a parabola in standard form with given focus and directrix.
- Identify the equation of an ellipse in standard form with given foci.
- Identify the equation of a hyperbola in standard form with given foci.
- Recognize a parabola, ellipse, or hyperbola from its eccentricity value.
- Write the polar equation of a conic section with eccentricity e.
- Identify when a general equation of degree two is a parabola, ellipse, or hyperbola.

Conic sections have been studied since the time of the ancient Greeks, and were considered to be an important mathematical concept. As early as 320 BCE, such Greek mathematicians as Menaechmus, Appollonius, and Archimedes were fascinated by these curves. Appollonius wrote an entire eight-volume treatise on conic sections in which he was, for example, able to derive a specific method for identifying a conic section through the use of geometry. Since then, important applications of conic sections have arisen (for example, in astronomy), and the properties of conic sections are used in radio telescopes, satellite dish receivers, and even architecture. In this section we discuss the three basic conic sections, some of their properties, and their equations.

Conic sections get their name because they can be generated by intersecting a plane with a cone. A cone has two identically shaped parts called **nappes.** One nappe is what most people mean by "cone," having the shape of a party hat. A right circular cone can be generated by revolving a line passing through the origin around the y-axis as shown in Figure 7.5.1.

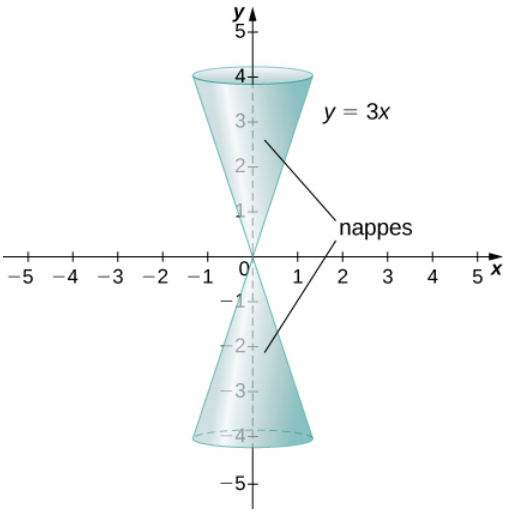


Figure 7.5.1: A cone generated by revolving the line $y = 3x$ around the y-axis.

Conic sections are generated by the intersection of a plane with a cone (Figure 7.5.2). If the plane is parallel to the axis of revolution (the y-axis), then the **conic section** is a hyperbola. If the plane is parallel to the generating line, the conic section is a parabola. If the plane is perpendicular to the axis of revolution, the conic section is a circle. If the plane intersects one nappe at an angle to the axis (other than **90°**), then the conic section is an ellipse.

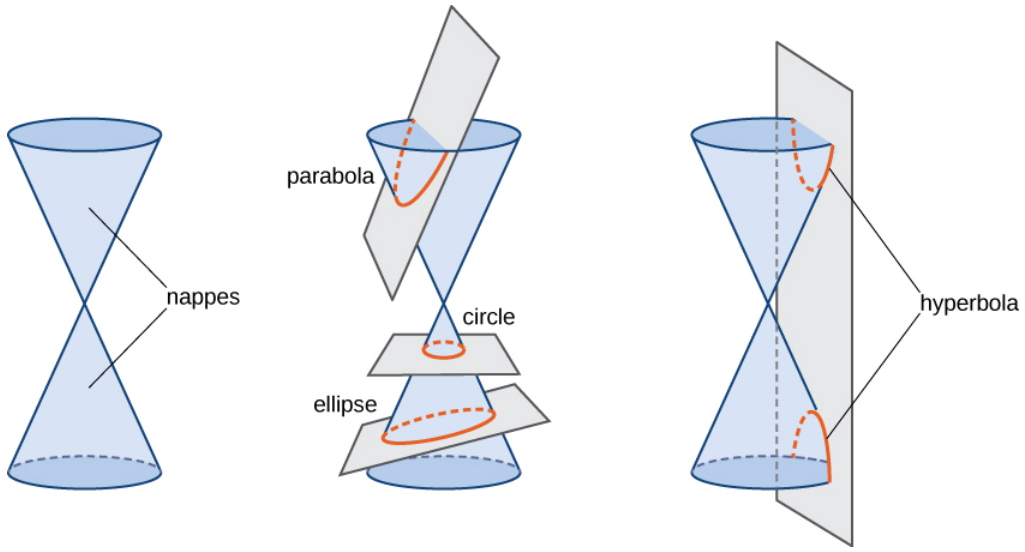


Figure 7.5.2: The four conic sections. Each conic is determined by the angle the plane makes with the axis of the cone.

Parabolas

A parabola is generated when a plane intersects a cone parallel to the generating line. In this case, the plane intersects only one of the nappes. A parabola can also be defined in terms of distances.

Definitions: The Focus, Directrix and Vertex

A parabola is the set of all points whose distance from a fixed point, called the *focus*, is equal to the distance from a fixed line, called the *directrix*. The point halfway between the focus and the directrix is called the *vertex* of the parabola.

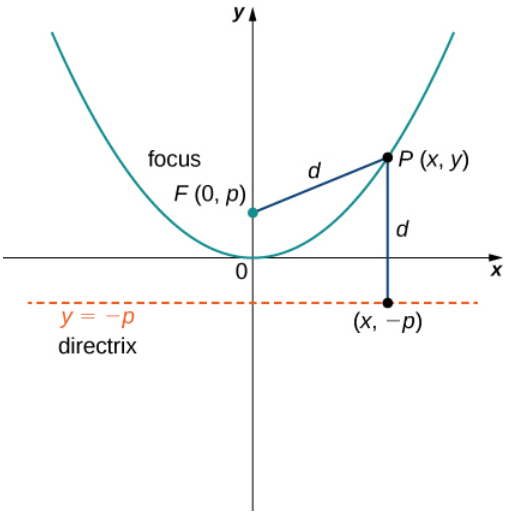


Figure 7.5.3: A typical parabola in which the distance from the focus to the vertex is represented by the variable p.

A graph of a typical parabola appears in Figure 7.5.3. Using this diagram in conjunction with the distance formula, we can derive an equation for a parabola. Recall the distance formula: Given point P with coordinates (x_1, y_1) and point Q with coordinates (x_2, y_2), the distance between them is given by the formula

$$d(P,Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\ .$$

Then from the definition of a parabola and Figure 7.5.3, we get

$$d(F,P) = d(P,Q)$$

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

Squaring both sides and simplifying yields

$$x^2+(p-y)^2 = 0^2+(-p-y)^2 \tag{7.5.1}$$
$$x^2+p^2-2py+y^2 = p^2+2py+y^2 \tag{7.5.2}$$
$$x^2-2py = 2py \tag{7.5.3}$$
$$x^2 = 4py. \tag{7.5.4}$$

Now suppose we want to relocate the vertex. We use the variables (h,k) to denote the coordinates of the vertex. Then if the focus is directly above the vertex, it has coordinates $(h,\ k+p)$ and the directrix has the equation $y=k-p$. Going through the same derivation yields the formula $(x-h)^2=4p(y-k)$. Solving this equation for y leads to the following theorem.

Equations for Parabolas: standard form

Given a parabola opening upward with vertex located at (h,k) and focus located at $(h,k+p)$, where p is a constant, the equation for the parabola is given by

$$y = \frac{1}{4p}(x-h)^2+k. \tag{7.5.5}$$

This is the *standard form* of a parabola.

We can also study the cases when the parabola opens down or to the left or the right. The equation for each of these cases can also be written in standard form as shown in the following graphs.

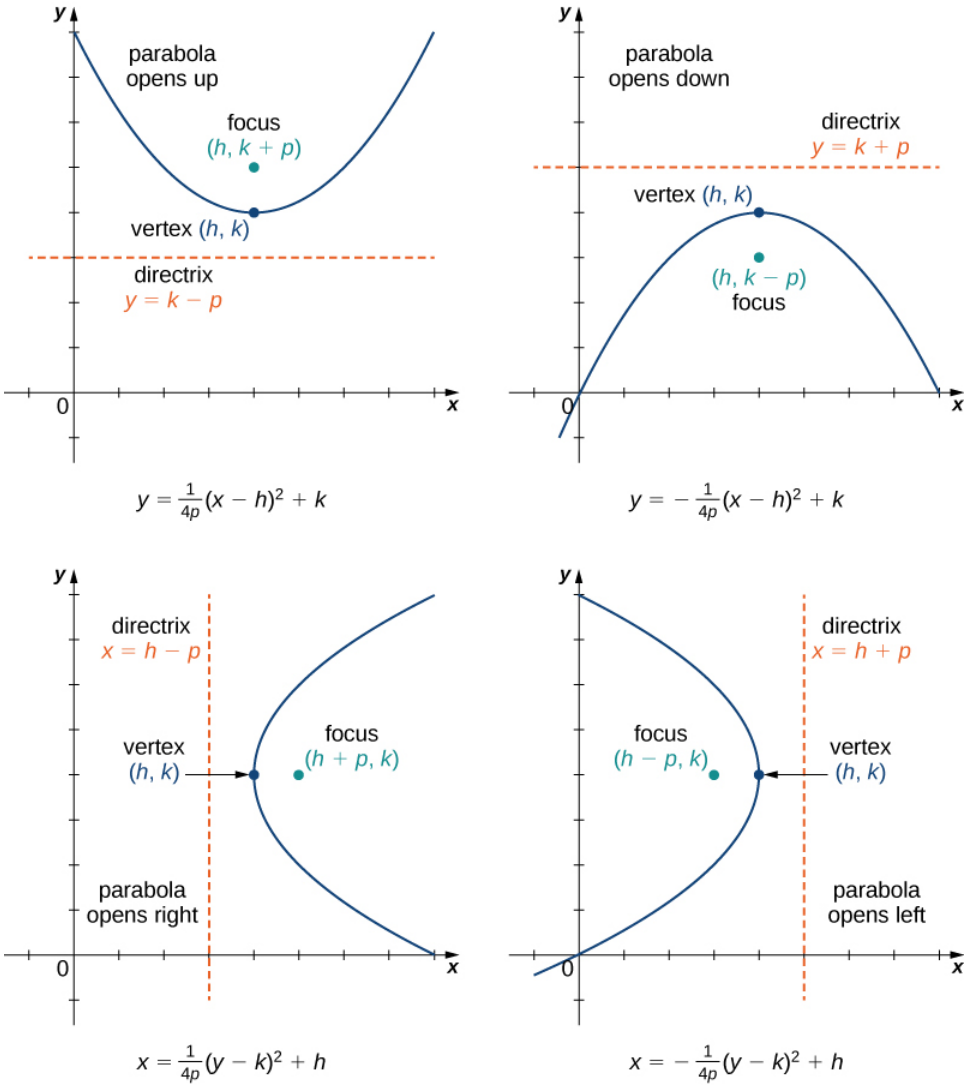


Figure 7.5.4: Four parabolas, opening in various directions, along with their equations in standard form.

In addition, the equation of a parabola can be written in the **general form**, though in this form the values of h, k, and p are not immediately recognizable. The general form of a parabola is written as

$$ax^2 + bx + cy + d = 0 \tag{7.5.6}$$

or

$$ay^2 + bx + cy + d = 0. \tag{7.5.7}$$

Equation 7.5.6 represents a parabola that opens either up or down. Equation 7.5.7 represents a parabola that opens either to the left or to the right. To put the equation into standard form, use the method of completing the square.

Example 7.5.1: Converting the Equation of a Parabola from General into Standard Form

Put the equation

$$x^2 - 4x - 8y + 12 = 0$$

into standard form and graph the resulting parabola.

Solution

Since y is not squared in this equation, we know that the parabola opens either upward or downward. Therefore we need to solve this equation for y, which will put the equation into standard form. To do that, first add $8y$ to both sides of the equation:

$$8y = x^2 - 4x + 12.$$

The next step is to complete the square on the right-hand side. Start by grouping the first two terms on the right-hand side using parentheses:

$$8y = (x^2 - 4x) + 12.$$

Next determine the constant that, when added inside the parentheses, makes the quantity inside the parentheses a perfect square trinomial. To do this, take half the coefficient of x and square it. This gives $\left(\frac{-4}{2}\right)^2 = 4$. Add 4 inside the parentheses and subtract 4 outside the parentheses, so the value of the equation is not changed:

$$8y = (x^2 - 4x + 4) + 12 - 4.$$

Now combine like terms and factor the quantity inside the parentheses:

$$8y = (x - 2)^2 + 8.$$

Finally, divide by 8:

$$y = \frac{1}{8}(x - 2)^2 + 1.$$

This equation is now in standard form. Comparing this to Equation 7.5.5 gives $h = 2$, $k = 1$, and $p = 2$. The parabola opens up, with vertex at $(2, 1)$, focus at $(2, 3)$, and directrix $y = -1$. The graph of this parabola appears as follows.

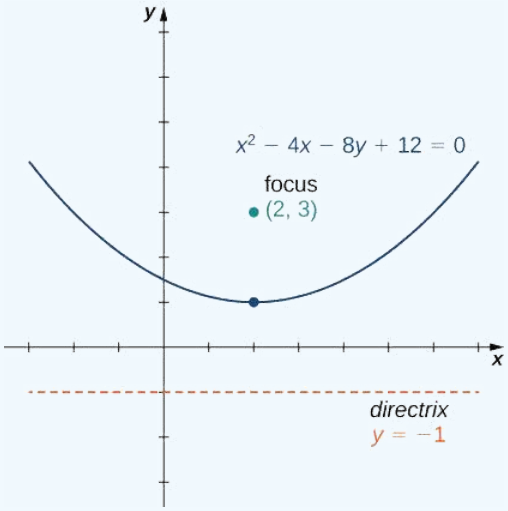


Figure 7.5.5: The parabola in Example 7.5.1.

Exercise 7.5.1

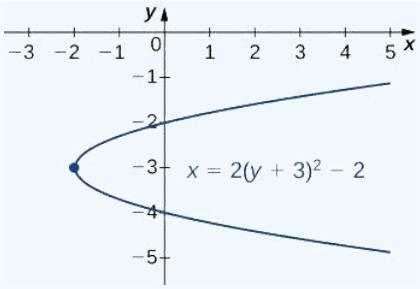
Put the equation $2y^2 - x + 12y + 16 = 0$ into standard form and graph the resulting parabola.

Hint

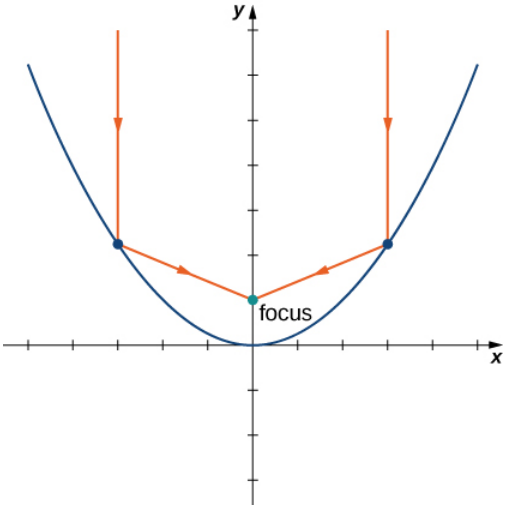
Solve for x. Check which direction the parabola opens.

Answer

$$x = 2(y + 3)^2 - 2$$



The axis of symmetry of a vertical (opening up or down) parabola is a vertical line passing through the vertex. The parabola has an interesting reflective property. Suppose we have a satellite dish with a parabolic cross section. If a beam of electromagnetic waves, such as light or radio waves, comes into the dish in a straight line from a satellite (parallel to the axis of symmetry), then the waves reflect off the dish and collect at the focus of the parabola as shown.



Consider a parabolic dish designed to collect signals from a satellite in space. The dish is aimed directly at the satellite, and a receiver is located at the focus of the parabola. Radio waves coming in from the satellite are reflected off the surface of the parabola to the receiver, which collects and decodes the digital signals. This allows a small receiver to gather signals from a wide angle of sky. Flashlights and headlights in a car work on the same principle, but in reverse: the source of the light (that is, the light bulb) is located at the focus and the reflecting surface on the parabolic mirror focuses the beam straight ahead. This allows a small light bulb to illuminate a wide angle of space in front of the flashlight or car.

Ellipses

An ellipse can also be defined in terms of distances. In the case of an ellipse, there are two foci (plural of focus), and two directrices (plural of directrix). We look at the directrices in more detail later in this section.

Definition: Ellipse

An ellipse is the set of all points for which the sum of their distances from two fixed points (the foci) is constant.

A graph of a typical ellipse is shown in Figure 7.5.6. In this figure the foci are labeled as F and F'. Both are the same fixed distance from the origin, and this distance is represented by the variable c. Therefore the coordinates of F are $(c, 0)$ and the coordinates of F' are $(-c, 0)$. The points P and P' are located at the ends of the **major axis** of the ellipse, and have coordinates $(a, 0)$ and $(-a, 0)$, respectively. The major axis is always the longest distance across the ellipse, and can be horizontal or vertical. Thus, the length of the major axis in this ellipse is $2a$. Furthermore, P and P' are called the vertices of the ellipse. The points Q and Q' are located at the ends of the **minor axis** of the ellipse, and have coordinates $(0, b)$ and $(0, -b)$, respectively. The minor axis is the shortest distance across the ellipse. The minor axis is perpendicular to the major axis.

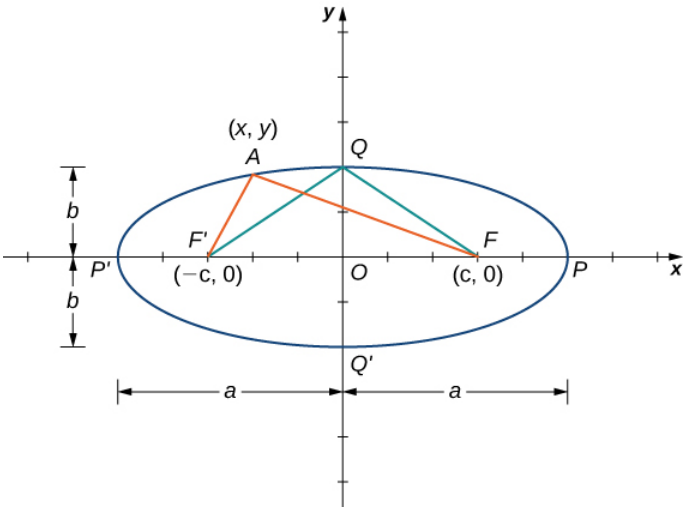


Figure 7.5.6: A typical ellipse in which the sum of the distances from any point on the ellipse to the foci is constant.

According to the definition of the ellipse, we can choose any point on the ellipse and the sum of the distances from this point to the two foci is constant. Suppose we choose the point P. Since the coordinates of point P are $(a, 0)$, the sum of the distances is

$$d(P, F) + d(P, F') = (a - c) + (a + c) = 2a.$$

Therefore the sum of the distances from an arbitrary point A with coordinates (x, y) is also equal to $2a$. Using the distance formula, we get

$$d(A, F) + d(A, F') = 2a.$$

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

Subtract the second radical from both sides and square both sides:

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

$$(x-c)^2 + y^2 = 4a^2 - 4a\sqrt{(x+c)^2 + y^2} + (x+c)^2 + y^2$$

$$x^2 - 2cx + c^2 + y^2 = 4a^2 - 4a\sqrt{(x+c)^2 + y^2} + x^2 + 2cx + c^2 + y^2$$

$$-2cx = 4a^2 - 4a\sqrt{(x+c)^2 + y^2} + 2cx.$$

Now isolate the radical on the right-hand side and square again:

$$-2cx = 4a^2 - 4a\sqrt{(x+c)^2 + y^2} + 2cx$$

$$4a\sqrt{(x+c)^2 + y^2} = 4a^2 + 4cx$$

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

$$(x+c)^2 + y^2 = a^2 + 2cx + \frac{c^2x^2}{a^2}$$

$$x^2 + 2cx + c^2 + y^2 = a^2 + 2cx + \frac{c^2x^2}{a^2}$$

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

Isolate the variables on the left-hand side of the equation and the constants on the right-hand side:

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

Divide both sides by $a^2 - c^2$. This gives the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1.$$

If we refer back to Figure 7.5.6, then the length of each of the two green line segments is equal to a. This is true because the sum of the distances from the point Q to the foci F and F' is equal to $2a$, and the lengths of these two line segments are equal. This line segment forms a right triangle with hypotenuse length a and leg lengths b and c. From the Pythagorean theorem, $b^2 + c^2 = a^2$ and $b^2 = a^2 - c^2$. Therefore the equation of the ellipse becomes

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Finally, if the center of the ellipse is moved from the origin to a point (h, k), we have the following standard form of an ellipse.

Equation of an Ellipse in Standard Form

Consider the ellipse with center (h, k), a horizontal major axis with length $2a$, and a vertical minor axis with length $2b$. Then the equation of this ellipse in standard form is

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1 \tag{7.5.8}$$

and the foci are located at $(h \pm c, k)$, where $c^2 = a^2 - b^2$. The equations of the directrices are $x = h \pm \frac{a^2}{c}$.

If the major axis is vertical, then the equation of the ellipse becomes

$$\frac{(x-h)^2}{b^2} + \frac{(y-k)^2}{a^2} = 1 \tag{7.5.9}$$

and the foci are located at $(h, k \pm c)$, where $c^2 = a^2 - b^2$. The equations of the directrices in this case are $y = k \pm \frac{a^2}{c}$.

If the major axis is horizontal, then the ellipse is called horizontal, and if the major axis is vertical, then the ellipse is called vertical. The equation of an ellipse is in general form if it is in the form

$$Ax^2 + By^2 + Cx + Dy + E = 0,$$

where A and B are either both positive or both negative. To convert the equation from general to standard form, use the method of **completing the square**.

Example 7.5.2: Finding the Standard Form of an Ellipse

Put the equation

$$9x^2 + 4y^2 - 36x + 24y + 36 = 0$$

into standard form and graph the resulting ellipse.

Solution

First subtract 36 from both sides of the equation:

$$9x^2 + 4y^2 - 36x + 24y = -36.$$

Next group the x terms together and the y terms together, and factor out the common factor:

$$(9x^2-36x)+(4y^2+24y)=-36$$

$$9(x^2-4x)+4(y^2+6y)=-36.$$

We need to determine the constant that, when added inside each set of parentheses, results in a perfect square. In the first set of parentheses, take half the coefficient of x and square it. This gives $\left(\frac{-4}{2}\right)^2=4$. In the second set of parentheses, take half the coefficient of y and square it. This gives $\left(\frac{6}{2}\right)^2=9$. Add these inside each pair of parentheses. Since the first set of parentheses has a 9 in front, we are actually adding 36 to the left-hand side. Similarly, we are adding 36 to the second set as well. Therefore the equation becomes

$$9(x^2-4x+4)+4(y^2+6y+9)=-36+36+36$$

$$9(x^2-4x+4)+4(y^2+6y+9)=36.$$

Now factor both sets of parentheses and divide by 36:

$$9(x-2)^2+4(y+3)^2=36$$

$$\frac{9(x-2)^2}{36}+\frac{4(y+3)^2}{36}=1$$

$$\frac{(x-2)^2}{4}+\frac{(y+3)^2}{9}=1.$$

The equation is now in standard form. Comparing this to Equation 7.5.9 gives $h=2, k=-3, a=3,$ and $b=2$. This is a vertical ellipse with center at $(2,-3)$, major axis 6, and minor axis 4. The graph of this ellipse appears as follows.

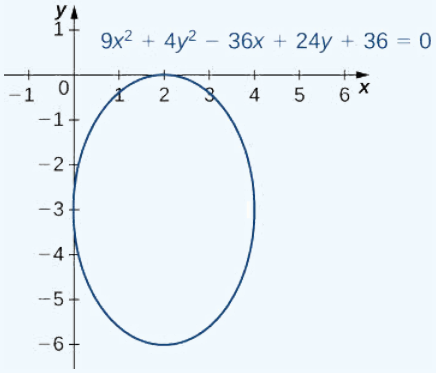


Figure 7.5.7: The ellipse in Example 7.5.2.

Exercise 7.5.2

Put the equation

$$9x^2+16y^2+18x-64y-71=0$$

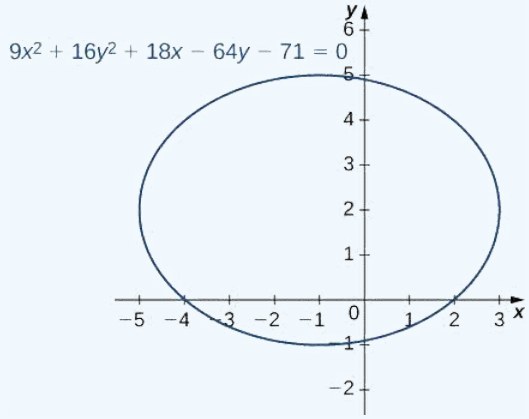
into standard form and graph the resulting ellipse.

Hint

Move the constant over and complete the square.

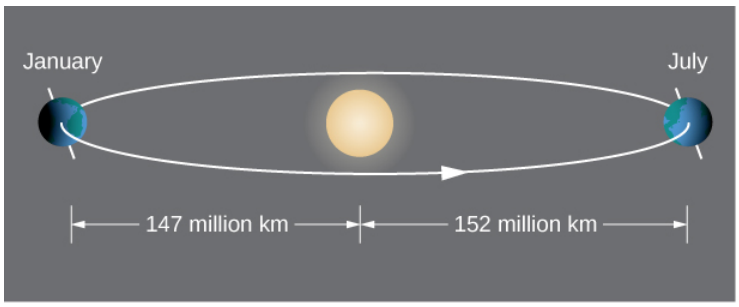
Answer

$$\frac{(x+1)^2}{16}+\frac{(y-2)^2}{9}=1$$



According to Kepler's first law of planetary motion, the orbit of a planet around the Sun is an ellipse with the Sun at one of the foci as shown in Figure 7.5.8*A*. Because Earth's orbit is an ellipse, the distance from the Sun varies throughout the year. A commonly held misconception is that Earth is closer to the Sun in the summer. In fact, in summer for the northern hemisphere, Earth is farther from the Sun than during winter. The difference in season is caused by the tilt of Earth's axis in the orbital plane. Comets that orbit the Sun, such as Halley's Comet, also have elliptical orbits, as do moons orbiting the planets and satellites orbiting Earth.

Ellipses also have interesting reflective properties: A light ray emanating from one focus passes through the other focus after mirror reflection in the ellipse. The same thing occurs with a sound wave as well. The National Statuary Hall in the U.S. Capitol in Washington, DC, is a famous room in an elliptical shape as shown in Figure 7.5.8*B*. This hall served as the meeting place for the U.S. House of Representatives for almost fifty years. The location of the two foci of this semi-elliptical room are clearly identified by marks on the floor, and even if the room is full of visitors, when two people stand on these spots and speak to each other, they can hear each other much more clearly than they can hear someone standing close by. Legend has it that John Quincy Adams had his desk located on one of the foci and was able to eavesdrop on everyone else in the House without ever needing to stand. Although this makes a good story, it is unlikely to be true, because the original ceiling produced so many echoes that the entire room had to be hung with carpets to dampen the noise. The ceiling was rebuilt in 1902 and only then did the now-famous whispering effect emerge. Another famous whispering gallery—the site of many marriage proposals—is in Grand Central Station in New York City.



(a)



(b)

Figure 7.5.8: (a) Earth's orbit around the Sun is an ellipse with the Sun at one focus. (b) Statuary Hall in the U.S. Capitol is a whispering gallery with an elliptical cross section.

Hyperbolas

A hyperbola can also be defined in terms of distances. In the case of a hyperbola, there are two foci and two directrices. Hyperbolas also have two asymptotes.

Definition: hyperbola

A hyperbola is the set of all points where the difference between their distances from two fixed points (the foci) is constant.

A graph of a typical hyperbola appears as follows.

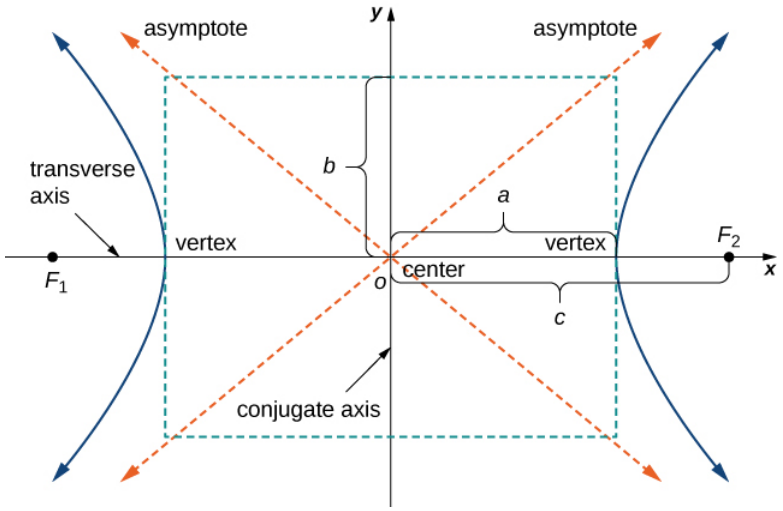


Figure 7.5.9: A typical hyperbola in which the difference of the distances from any point on the hyperbola to the foci is constant. The transverse axis is also called the major axis, and the conjugate axis is also called the minor axis.

The derivation of the equation of a hyperbola in standard form is virtually identical to that of an ellipse. One slight hitch lies in the definition: The difference between two numbers is always positive. Let P be a point on the hyperbola with coordinates (x, y). Then the definition of the hyperbola gives $|d(P, F_1) - d(P, F_2)| = constant$. To simplify the derivation, assume that P is on the right branch of the hyperbola, so the absolute value bars drop. If it is on the left branch, then the subtraction is reversed. The vertex of the right branch has coordinates $(a, 0)$, so

$$d(P, F_1) - d(P, F_2) = (c + a) - (c - a) = 2a.$$

This equation is therefore true for any point on the hyperbola. Returning to the coordinates (x, y) for P:

$$d(P, F_1) - d(P, F_2) = 2a$$

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

Isolate the second radical and square both sides:

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

$$(x - c)^2 + y^2 = 4a^2 - 4a\sqrt{(x + c)^2 + y^2} + (x + c)^2 + y^2$$

$$x^2 - 2cx + c^2 + y^2 = 4a^2 - 4a\sqrt{(x + c)^2 + y^2} + x^2 + 2cx + c^2 + y^2$$

$$-2cx = 4a^2 - 4a\sqrt{(x + c)^2 + y^2} + 2cx.$$

Now isolate the radical on the right-hand side and square again:

$$-2cx = 4a^2 - 4a\sqrt{(x + c)^2 + y^2} + 2cx$$

$$-4a\sqrt{(x + c)^2 + y^2} = -4a^2 - 4cx$$

$$-\sqrt{(x + c)^2 + y^2} = -a - \frac{cx}{a}$$

$$(x + c)^2 + y^2 = a^2 + 2cx + \frac{c^2x^2}{a^2}$$

$$x^2 + 2cx + c^2 + y^2 = a^2 + 2cx + \frac{c^2x^2}{a^2}$$

$$x^2 + c^2 + y^2 = a^2 + \frac{c^2x^2}{a^2} \quad .$$

Isolate the variables on the left-hand side of the equation and the constants on the right-hand side:

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

Finally, divide both sides by $a^2 - c^2$. This gives the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1.$$

We now define b so that $b^2 = c^2 - a^2$. This is possible because $c > a$. Therefore the equation of the hyperbola becomes

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Finally, if the center of the hyperbola is moved from the origin to the point (h, k), we have the following standard form of a hyperbola.

Equation of a Hyperbola in Standard Form

Consider the hyperbola with center (h, k), a horizontal major axis, and a vertical minor axis. Then the equation of this hyperbola is

$$\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1 \tag{7.5.10}$$

and the foci are located at $(h \pm c, k)$, where $c^2 = a^2 + b^2$. The equations of the asymptotes are given by $y = k \pm \frac{b}{a}(x - h)$. The equations of the directrices are

$$x = h \pm \frac{a^2}{\sqrt{a^2 + b^2}} = h \pm \frac{a^2}{c}$$

If the major axis is vertical, then the equation of the hyperbola becomes

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

and the foci are located at $(h, k \pm c)$, where $c^2 = a^2 + b^2$. The equations of the asymptotes are given by $y = k \pm \frac{a}{b}(x - h)$. The equations of the directrices are

$$y = k \pm \frac{a^2}{\sqrt{a^2 + b^2}} = k \pm \frac{a^2}{c}.$$

If the major axis (transverse axis) is horizontal, then the hyperbola is called horizontal, and if the major axis is vertical then the hyperbola is called vertical. The equation of a hyperbola is in general form if it is in the form

$$Ax^2 + By^2 + Cx + Dy + E = 0,$$

where A and B have opposite signs. In order to convert the equation from general to standard form, use the method of completing the square.

Example 7.5.3: Finding the Standard Form of a Hyperbola

Put the equation $9x^2 - 16y^2 + 36x + 32y - 124 = 0$ into standard form and graph the resulting hyperbola. What are the equations of the asymptotes?

Solution

First add 124 to both sides of the equation:

$$9x^2 - 16y^2 + 36x + 32y = 124.$$

Next group the x terms together and the y terms together, then factor out the common factors:

$$(9x^2+36x)-(16y^2-32y)=124$$
$$9(x^2+4x)-16(y^2-2y)=124\ .$$

We need to determine the constant that, when added inside each set of parentheses, results in a perfect square. In the first set of parentheses, take half the coefficient of x and square it. This gives $\left(\frac{4}{2}\right)^2=4$. In the second set of parentheses, take half the coefficient of y and square it. This gives $\left(\frac{-2}{2}\right)^2=1$. Add these inside each pair of parentheses. Since the first set of parentheses has a 9 in front, we are actually adding 36 to the left-hand side. Similarly, we are subtracting 16 from the second set of parentheses. Therefore the equation becomes

$$9(x^2+4x+4)-16(y^2-2y+1)=124+36-16$$
$$9(x^2+4x+4)-16(y^2-2y+1)=144.$$

Next factor both sets of parentheses and divide by 144:

$$9(x+2)^2-16(y-1)^2=144$$
$$\frac{9(x+2)^2}{144}-\frac{16(y-1)^2}{144}=1$$
$$\frac{(x+2)^2}{16}-\frac{(y-1)^2}{9}=1.$$

The equation is now in standard form. Comparing this to Equation 7.5.10 gives $h=-2$, $k=1$, $a=4$, and $b=3$. This is a horizontal hyperbola with center at $(-2,1)$ and asymptotes given by the equations $y=1\pm\frac{3}{4}(x+2)$. The graph of this hyperbola appears in Figure 7.5.10.

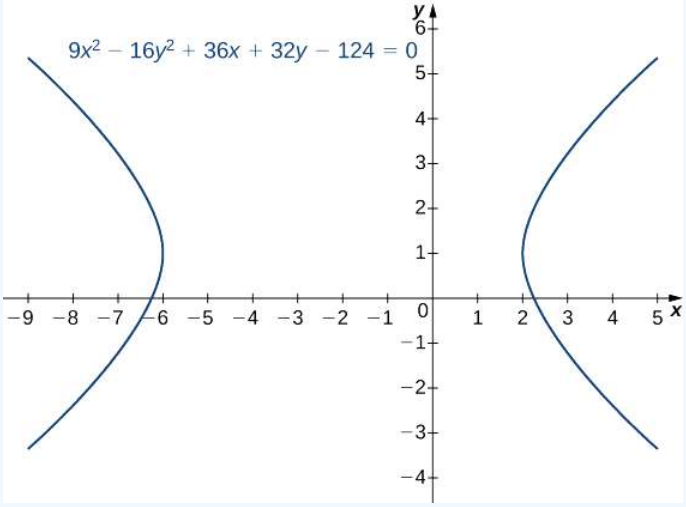


Figure 7.5.10: Graph of the hyperbola in Example 7.5.3.

? Exercise 7.5.3

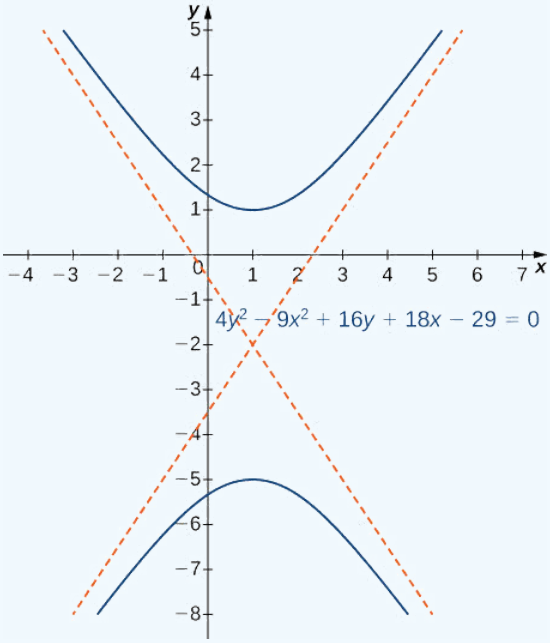
Put the equation $4y^2-9x^2+16y+18x-29=0$ into standard form and graph the resulting hyperbola. What are the equations of the asymptotes?

Hint

Move the constant over and complete the square. Check which direction the hyperbola opens

Answer

$\frac{(y+2)^2}{9} - \frac{(x-1)^2}{4} = 1$. This is a vertical hyperbola. Asymptotes $y = -2 \pm \frac{3}{2}(x-1)$.



Hyperbolas also have interesting reflective properties. A ray directed toward one focus of a hyperbola is reflected by a hyperbolic mirror toward the other focus. This concept is illustrated in Figure 7.5.11.

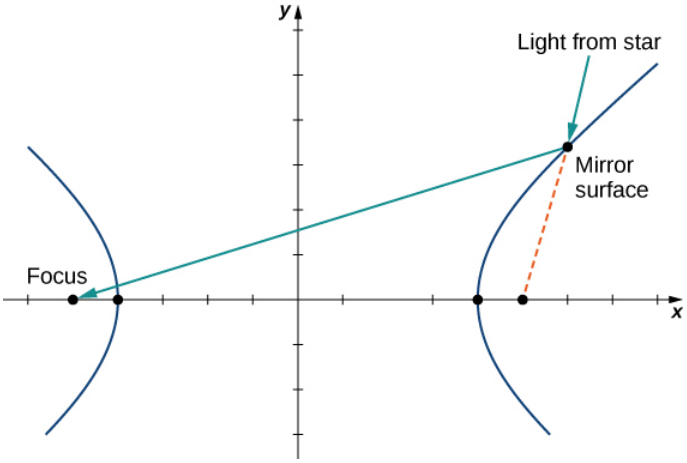


Figure 7.5.11: A hyperbolic mirror used to collect light from distant stars.

This property of the hyperbola has important applications. It is used in radio direction finding (since the difference in signals from two towers is constant along hyperbolas), and in the construction of mirrors inside telescopes (to reflect light coming from the parabolic mirror to the eyepiece). Another interesting fact about hyperbolas is that for a comet entering the solar system, if the speed is great enough to escape the Sun's gravitational pull, then the path that the comet takes as it passes through the solar system is hyperbolic.

Eccentricity and Directrix

An alternative way to describe a conic section involves the directrices, the foci, and a new property called eccentricity. We will see that the value of the eccentricity of a conic section can uniquely define that conic.

Definition: Eccentricity and Directrices

The *eccentricity* e of a conic section is defined to be the distance from any point on the conic section to its focus, divided by the perpendicular distance from that point to the nearest directrix. This value is constant for any conic section, and can define the conic section as well:

1. If $e = 1$, the conic is a parabola.
2. If $e < 1$, it is an ellipse.
3. If $e > 1$, it is a hyperbola.

The eccentricity of a circle is zero. The *directrix* of a conic section is the line that, together with the point known as the focus, serves to define a conic section. Hyperbolas and noncircular ellipses have two foci and two associated directrices. Parabolas have one focus and one directrix.

The three conic sections with their directrices appear in Figure 7.5.12

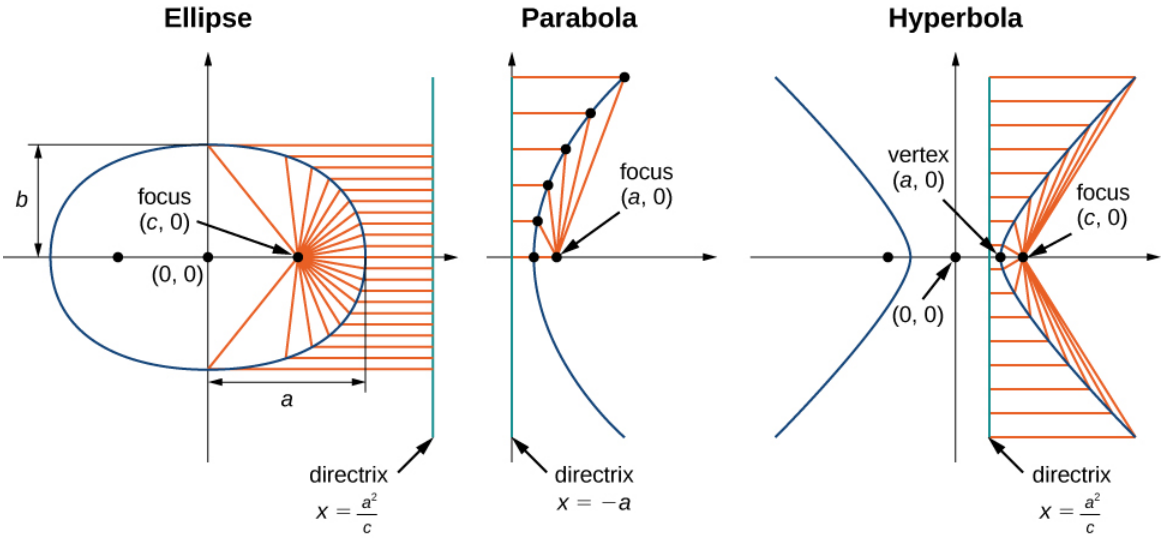


Figure 7.5.12: The three conic sections with their foci and directrices.

Recall from the definition of a parabola that the distance from any point on the parabola to the focus is equal to the distance from that same point to the directrix. Therefore, by definition, the eccentricity of a parabola must be 1. The equations of the directrices of a horizontal ellipse are $x = \pm\dfrac{a^2}{c}$. The right vertex of the ellipse is located at $(a, 0)$ and the right focus is $(c, 0)$. Therefore the distance from the vertex to the focus is $a - c$ and the distance from the vertex to the right directrix is $\dfrac{a^2}{c} - c$. This gives the eccentricity as

$$e = \frac{a-c}{\frac{a^2}{c} - a} = \frac{c(a-c)}{a^2 - ac} = \frac{c(a-c)}{a(a-c)} = \frac{c}{a}.$$

Since $c < a$, this step proves that the eccentricity of an ellipse is less than 1. The directrices of a horizontal hyperbola are also located at $x = \pm\dfrac{a^2}{c}$, and a similar calculation shows that the eccentricity of a hyperbola is also $e = \dfrac{c}{a}$. However in this case we have $c > a$, so the eccentricity of a hyperbola is greater than 1.

Example 7.5.4: Determining Eccentricity of a Conic Section

Determine the eccentricity of the ellipse described by the equation

$$\frac{(x-3)^2}{16} + \frac{(y+2)^2}{25} = 1.$$

Solution

From the equation we see that $a=5$ and $b=4$. The value of c can be calculated using the equation $a^2=b^2+c^2$ for an ellipse. Substituting the values of a and b and solving for c gives $c=3$. Therefore the eccentricity of the ellipse is $e=\dfrac{c}{a}=\dfrac{3}{5}=0.6$.

Exercise 7.5.4

Determine the eccentricity of the hyperbola described by the equation

$$\frac{(y-3)^2}{49}-\frac{(x+2)^2}{25}=1.$$

Hint

First find the values of a and b, then determine c using the equation $c^2=a^2+b^2$.

Answer

$$e=\frac{c}{a}=\frac{\sqrt{74}}{7}\approx 1.229$$

Polar Equations of Conic Sections

Sometimes it is useful to write or identify the equation of a conic section in polar form. To do this, we need the concept of the focal parameter. The **focal parameter** of a conic section p is defined as the distance from a focus to the nearest directrix. The following table gives the focal parameters for the different types of conics, where a is the length of the semi-major axis (i.e., half the length of the major axis), c is the distance from the origin to the focus, and e is the eccentricity. In the case of a parabola, a represents the distance from the vertex to the focus.

Table 7.5.1: Eccentricities and Focal Parameters of the Conic Sections

Conic	e	p
Ellipse	$0<e<1$	$\dfrac{a^2-c^2}{c}=\dfrac{a(1-e^2)}{c}$
Parabola	$e=1$	$2a$
Hyperbola	$e>1$	$\dfrac{c^2-a^2}{c}=\dfrac{a(e^2-1)}{c}$

Using the definitions of the focal parameter and eccentricity of the conic section, we can derive an equation for any conic section in polar coordinates. In particular, we assume that one of the foci of a given conic section lies at the pole. Then using the definition of the various conic sections in terms of distances, it is possible to prove the following theorem.

Polar Equation of Conic Sections

The polar equation of a conic section with focal parameter p is given by

$$r=\frac{ep}{1\pm e\cos\theta}\quad\text{or}\quad r=\frac{ep}{1\pm e\sin\theta}.$$

In the equation on the left, the major axis of the conic section is horizontal, and in the equation on the right, the major axis is vertical. To work with a conic section written in polar form, first make the constant term in the denominator equal to 1. This can be done by dividing both the numerator and the denominator of the fraction by the constant that appears in front of the plus or minus in the denominator. Then the coefficient of the sine or cosine in the denominator is the eccentricity. This value identifies the conic. If cosine appears in the denominator, then the conic is horizontal. If sine appears, then the conic is vertical. If both appear then the axes are rotated. The center of the conic is not necessarily at the origin. The center is at the origin only if the conic is a circle (i.e., $e=0$).

Example 7.5.5: Graphing a Conic Section in Polar Coordinates

Identify and create a graph of the conic section described by the equation

$$r = \frac{3}{1+2\cos\theta}.$$

Solution

The constant term in the denominator is 1, so the eccentricity of the conic is 2. This is a hyperbola. The focal parameter p can be calculated by using the equation $ep = 3$. Since $e = 2$, this gives $p = \frac{3}{2}$. The cosine function appears in the denominator, so the hyperbola is horizontal. Pick a few values for θ and create a table of values. Then we can graph the hyperbola (Figure 7.5.13).

θ	r	θ	r
0	1	π	-3
$\frac{\pi}{4}$	$\frac{3}{1+\sqrt{2}} \approx 1.2426$	$\frac{5\pi}{4}$	$\frac{3}{1-\sqrt{2}} \approx -7.2426$
$\frac{\pi}{2}$	3	$\frac{3\pi}{2}$	3
$\frac{3\pi}{4}$	$\frac{3}{1-\sqrt{2}} \approx -7.2426$	$\frac{7\pi}{4}$	$\frac{3}{1+\sqrt{2}} \approx 1.2426$

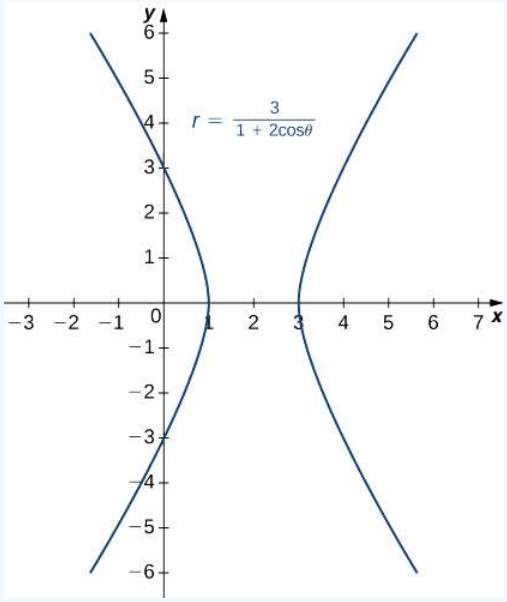


Figure 7.5.13: Graph of the hyperbola described in Example 7.5.5.

Exercise 7.5.5

Identify and create a graph of the conic section described by the equation

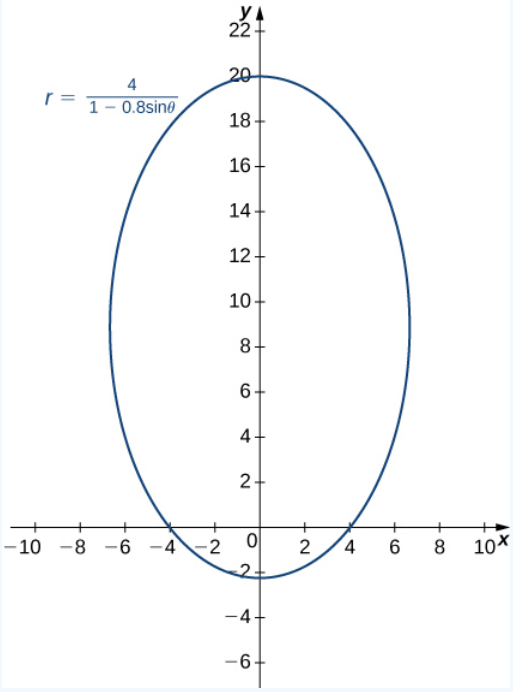
$$r = \frac{4}{1-0.8\sin\theta}.$$

Hint

First find the values of e and p, and then create a table of values.

Answer

Here $e = 0.8$ and $p = 5$. This conic section is an ellipse.



General Equations of Degree Two

A general equation of degree two can be written in the form

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

The graph of an equation of this form is a conic section. If $B \neq 0$ then the coordinate axes are rotated. To identify the conic section, we use the discriminant of the conic section $4AC - B^2$.

Identifying the Conic Section

One of the following cases must be true:

1. $4AC - B^2 > 0$. If so, the graph is an ellipse.
2. $4AC - B^2 = 0$. If so, the graph is a parabola.
3. $4AC - B^2 < 0$. If so, the graph is a hyperbola.

The simplest example of a second-degree equation involving a cross term is $xy = 1$. This equation can be solved for y to obtain $y = \dfrac{1}{x}$. The graph of this function is called a rectangular hyperbola as shown.

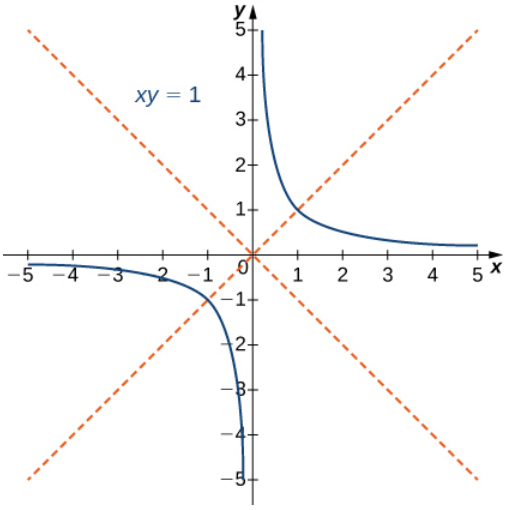


Figure 7.5.14: Graph of the equation $xy = 1$; The red lines indicate the rotated axes.

The asymptotes of this hyperbola are the x and y coordinate axes. To determine the angle θ of rotation of the conic section, we use the formula $\cot 2\theta = \frac{A-C}{B}$. In this case $A = C = 0$ and $B = 1$, so $\cot 2\theta = (0-0)/1 = 0$ and $\theta = 45^\circ$. The method for graphing a conic section with rotated axes involves determining the coefficients of the conic in the rotated coordinate system. The new coefficients are labeled A', B', C', D', E',and F', and are given by the formulas

$$\begin{aligned} A' &= A\cos^2\theta + B\cos\theta\sin\theta + C\sin^2\theta \\ B' &= 0 \\ C' &= A\sin^2\theta - B\sin\theta\cos\theta + C\cos^2\theta \\ D' &= D\cos\theta + E\sin\theta \\ E' &= -D\sin\theta + E\cos\theta \\ F' &= F. \end{aligned}$$

Procedure: graphing a rotated conic

The procedure for graphing a rotated conic is the following:

1. Identify the conic section using the discriminant $4AC - B^2$.
2. Determine θ using the formula

$$\cot 2\theta = \frac{A-C}{B}. \tag{7.5.11}$$

3. Calculate A', B', C', D', E', and F'.
4. Rewrite the original equation using A', B', C', D', E', and F'.
5. Draw a graph using the rotated equation.

Example 7.5.6: Identifying a Rotated Conic

Identify the conic and calculate the angle of rotation of axes for the curve described by the equation

$$13x^2 - 6\sqrt{3}xy + 7y^2 - 256 = 0.$$

Solution

In this equation, $A = 13$, $B = -6\sqrt{3}$, $C = 7$, $D = 0$, $E = 0$, and $F = -256$. The discriminant of this equation is

$$4AC - B^2 = 4(13)(7) - (-6\sqrt{3})^2 = 364 - 108 = 256.$$

Therefore this conic is an ellipse.

To calculate the angle of rotation of the axes, use Equation 7.5.11

$$\cot 2\theta = \frac{A-C}{B}.$$

This gives

$$\cot 2\theta = \frac{A-C}{B} = \frac{13-7}{-6\sqrt{3}} = -\frac{\sqrt{3}}{3}\ .$$

Therefore $2\theta = 120^o$ and $\theta = 60^o$, which is the angle of the rotation of the axes.

To determine the rotated coefficients, use the formulas given above:

$$\begin{aligned}
A' &= A\cos^2\theta + B\cos\theta\sin\theta + C\sin^2\theta \\
&= 13\cos^2 60 + (-6\sqrt{3})\cos 60\sin 60 + 7\sin^2 60 \\
&= 13\left(\frac{1}{2}\right)^2 - 6\sqrt{3}\left(\frac{1}{2}\right)\left(\frac{\sqrt{3}}{2}\right) + 7\left(\frac{\sqrt{3}}{2}\right)^2 \\
&= 4, \\
B' &= 0 \\
C' &= A\sin^2\theta - B\sin\theta\cos\theta + C\cos^2\theta \\
&= 13\sin^2 60 + (6\sqrt{3})\sin 60\cos 60 + 7\cos^2 60 \\
&= 13\left(\frac{\sqrt{3}}{2}\right)^2 + 6\sqrt{3}\left(\frac{\sqrt{3}}{2}\right)\left(\frac{1}{2}\right) + 7\left(\frac{1}{2}\right)^2 \\
&= 16, \\
D' &= D\cos\theta + E\sin\theta \\
&= (0)\cos 60 + (0)\sin 60 \\
&= 0, \\
E' &= -D\sin\theta + E\cos\theta \\
&= -(0)\sin 60 + (0)\cos 60 \\
&= 0 \\
F' &= F \\
&= -256.
\end{aligned}$$

The equation of the conic in the rotated coordinate system becomes

$$\begin{aligned}
4(x')^2 + 16(y')^2 &= 256 \\
\frac{(x')^2}{64} + \frac{(y')^2}{16} &= 1\ .
\end{aligned}$$

A graph of this conic section appears as follows.

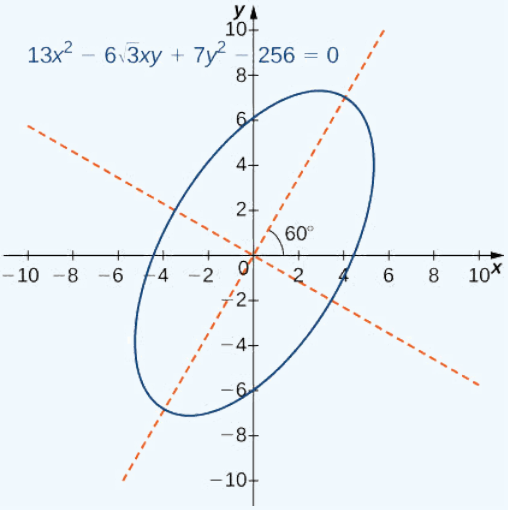


Figure 7.5.15: Graph of the ellipse described by the equation $13x^2 - 6\sqrt{3}xy + 7y^2 - 256 = 0$. The axes are rotated 60°. The red dashed lines indicate the rotated axes.

? Exercise 7.5.6

Identify the conic and calculate the angle of rotation of axes for the curve described by the equation

$$3x^2 + 5xy - 2y^2 - 125 = 0.$$

Hint

Follow steps 1 and 2 of the five-step method outlined above

Answer

The conic is a hyperbola and the angle of rotation of the axes is $\theta = 22.5^\circ$.

Key Concepts

- The equation of a vertical parabola in standard form with given focus and directrix is $y = \frac{1}{4p}(x-h)^2 + k$ where p is the distance from the vertex to the focus and (h, k) are the coordinates of the vertex.
- The equation of a horizontal ellipse in standard form is $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ where the center has coordinates (h, k), the major axis has length $2a$, the minor axis has length $2b$, and the coordinates of the foci are $(h \pm c, k)$, where $c^2 = a^2 - b^2$.
- The equation of a horizontal hyperbola in standard form is $\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1$ where the center has coordinates (h, k), the vertices are located at $(h \pm a, k)$, and the coordinates of the foci are $(h \pm c, k)$, where $c^2 = a^2 + b^2$.
- The eccentricity of an ellipse is less than 1, the eccentricity of a parabola is equal to 1, and the eccentricity of a hyperbola is greater than 1. The eccentricity of a circle is 0.
- The polar equation of a conic section with eccentricity e is $r = \frac{ep}{1 \pm e\cos\theta}$ or $r = \frac{ep}{1 \pm e\sin\theta}$, where p represents the focal parameter.
- To identify a conic generated by the equation $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$,first calculate the discriminant $D = 4AC - B^2$. If $D > 0$ then the conic is an ellipse, if $D = 0$ then the conic is a parabola, and if $D < 0$ then the conic is a hyperbola.

Glossary

conic section

a conic section is any curve formed by the intersection of a plane with a cone of two nappes

directrix

a directrix (plural: directrices) is a line used to construct and define a conic section; a parabola has one directrix; ellipses and hyperbolas have two

discriminant

the value $4AC - B^2$, which is used to identify a conic when the equation contains a term involving xy, is called a discriminant

focus

a focus (plural: foci) is a point used to construct and define a conic section; a parabola has one focus; an ellipse and a hyperbola have two

eccentricity

the eccentricity is defined as the distance from any point on the conic section to its focus divided by the perpendicular distance from that point to the nearest directrix

focal parameter

the focal parameter is the distance from a focus of a conic section to the nearest directrix

general form

an equation of a conic section written as a general second-degree equation

major axis

the major axis of a conic section passes through the vertex in the case of a parabola or through the two vertices in the case of an ellipse or hyperbola; it is also an axis of symmetry of the conic; also called the transverse axis

minor axis

the minor axis is perpendicular to the major axis and intersects the major axis at the center of the conic, or at the vertex in the case of the parabola; also called the conjugate axis

nappe

a nappe is one half of a double cone

standard form

an equation of a conic section showing its properties, such as location of the vertex or lengths of major and minor axes

vertex

a vertex is an extreme point on a conic section; a parabola has one vertex at its turning point. An ellipse has two vertices, one at each end of the major axis; a hyperbola has two vertices, one at the turning point of each branch

---

This page titled 7.5: Conic Sections is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **11.5: Conic Sections** by Edwin “Jed” Herman, Gilbert Strang is licensed CC BY-NC-SA 4.0. Original source: https://openstax.org/details/books/calculus-volume-1.

Index

A

absolute convergence
5.5: Alternating Series
absolute error
3.6: Numerical Integration
Airy's equation
6.4: Working with Taylor Series
Alternating Harmonic Series
5.5: Alternating Series
alternating series
5.5: Alternating Series
Alternating Series Test
5.5: Alternating Series
angular coordinate
7.3: Polar Coordinates
arc length
2.4: Arc Length of a Curve and Surface Area
Arc length of a parametric curve
7.2: Calculus of Parametric Curves
Arc length of a polar curve
7.4: Area and Arc Length in Polar Coordinates
Archimedean spiral
7.3: Polar Coordinates
AREA BETWEEN TWO CURVES
2.1: Areas between Curves
Area of a region bounded by a polar curve
7.4: Area and Arc Length in Polar Coordinates
Area under a parametric curve
7.2: Calculus of Parametric Curves
Areas of Compound Regions
2.1: Areas between Curves
arithmetic sequence
5.1: Sequences
asymptotically stable solution
4.2: Direction Fields and Numerical Methods
asymptotically unstable solution
4.2: Direction Fields and Numerical Methods
autonomous differential equation
4.3: Separable Equations

B

binomial series
6.4: Working with Taylor Series

C

cardioid
7.3: Polar Coordinates
carrying capacity
4.4: The Logistic Equation
catenary
2.9: Calculus of the Hyperbolic Functions
center of mass
2.6: Moments and Centers of Mass
comparison test
5.4: Comparison Tests
compound interest
2.8: Exponential Growth and Decay
computer algebra system
3.5: Other Strategies for Integration
conditional convergence
5.5: Alternating Series
CONIC SECTIONS
7.5: Conic Sections
convergence of a series
5.2: Infinite Series
convergent sequence
5.1: Sequences
curtate cycloid
7.1: Parametric Equations
cusp
7.1: Parametric Equations
cusps
7.1: Parametric Equations
cycloid
7.1: Parametric Equations
Cycloids
7.1: Parametric Equations

D

DENSITY FUNCTION
2.5: Physical Applications of Integration
Derivative of Hyperbolic Functions
2.9: Calculus of the Hyperbolic Functions
Derivative of parametric equations
7.2: Calculus of Parametric Curves
differential equation
4.1: Basics of Differential Equations
directrix
7.5: Conic Sections
discriminant
7.5: Conic Sections
disk method
2.2: Determining Volumes by Slicing
divergence of a series
5.2: Infinite Series
Divergence Test
5.3: The Divergence and Integral Tests
divergent sequence
5.1: Sequences
doubling time
2.8: Exponential Growth and Decay

E

eccentricity
7.5: Conic Sections
ellipses
7.5: Conic Sections
equilibrium solution
4.2: Direction Fields and Numerical Methods
Euler's Method
4.2: Direction Fields and Numerical Methods
exponential function
2.7: Integrals, Exponential Functions, and Logarithms

F

family of solutions
4.1: Basics of Differential Equations
Fibonacci Numbers
5.1: Sequences
first moment
2.6: Moments and Centers of Mass
focal parameter
7.5: Conic Sections
focus
7.5: Conic Sections
frustum
2.4: Arc Length of a Curve and Surface Area

G

general form
7.5: Conic Sections
general solution
4.1: Basics of Differential Equations
Geometric Sequences
5.1: Sequences
geometric series
5.2: Infinite Series
5.5: Alternating Series
growth rate
4.4: The Logistic Equation

H

harmonic series
5.2: Infinite Series
Hooke's law
2.5: Physical Applications of Integration
hydrostatic pressure
2.5: Physical Applications of Integration
hyperbolas
7.5: Conic Sections
hypocycloid
7.1: Parametric Equations

I

improper integral
3.7: Improper Integrals
index variable
5.1: Sequences
infinite series
5.2: Infinite Series
initial population
4.4: The Logistic Equation
initial value
4.1: Basics of Differential Equations
initial velocity
4.1: Basics of Differential Equations
Integral Test
5.3: The Divergence and Integral Tests
integration by parts
3.1: Integration by Parts
integration table
3.5: Other Strategies for Integration
interval of convergence
6.1: Power Series and Functions

L

limacon
7.3: Polar Coordinates
Limit Comparison Test
5.4: Comparison Tests

limit of a sequence
5.1: Sequences
logistic differential equation
4.4: The Logistic Equation

M
Maclaurin polynomial
6.3: Taylor and Maclaurin Series
Maclaurin series
6.3: Taylor and Maclaurin Series
major axis
7.5: Conic Sections
method of cylindrical shells
2.3: Volumes of Revolution - Cylindrical Shells
method of equating coefficients
3.4: Partial Fractions
method of strategic substitution
3.4: Partial Fractions
Midpoint Rule
3.6: Numerical Integration
minor axis
7.5: Conic Sections
Moments of Inertia
2.6: Moments and Centers of Mass
Monotone Convergence Theorem
5.1: Sequences
monotone sequence
5.1: Sequences

N
nappe
7.5: Conic Sections
nappes
7.5: Conic Sections
Natural logarithm function
2.7: Integrals, Exponential Functions, and Logarithms
Newton's Law of Cooling
2.8: Exponential Growth and Decay
4.3: Separable Equations
nonelementary integral
6.4: Working with Taylor Series
numerical integration
3.6: Numerical Integration

O
order of a differential equation
4.1: Basics of Differential Equations
orientation
7.1: Parametric Equations

P
parabola
7.5: Conic Sections
parameter
7.1: Parametric Equations
parameterization of a curve
7.1: Parametric Equations
parametric curve
7.1: Parametric Equations
parametric equations
7.1: Parametric Equations
partial fraction decomposition
3.4: Partial Fractions
partial sum
5.2: Infinite Series
particular solution
4.1: Basics of Differential Equations
phase line
4.4: The Logistic Equation
polar axis
7.3: Polar Coordinates
polar coordinate system
7.3: Polar Coordinates
polar equation
7.3: Polar Coordinates
pole
7.3: Polar Coordinates
population growth
2.8: Exponential Growth and Decay
Power Reduction Formula
3.2: Trigonometric Integrals
power series
6: Power Series
6.1: Power Series and Functions

R
radial coordinate
7.3: Polar Coordinates
radial density
2.5: Physical Applications of Integration
radius of convergence
6.1: Power Series and Functions
ratio test
5.6: Ratio and Root Tests
recurrence relation
5.1: Sequences
relative error
3.6: Numerical Integration
Remainder Estimate
5.3: The Divergence and Integral Tests
repeated linear factors
3.4: Partial Fractions
Riemann Rearrangement Theorem
5.5: Alternating Series
root test
5.6: Ratio and Root Tests
rose
7.3: Polar Coordinates

S
separation of variables
4.3: Separable Equations
sequence
5.1: Sequences
sequences
5.1: Sequences
slicing method
2.2: Determining Volumes by Slicing
slope field
4.2: Direction Fields and Numerical Methods
solid of revolution
2.2: Determining Volumes by Slicing
2.3: Volumes of Revolution - Cylindrical Shells
solution curve
4.2: Direction Fields and Numerical Methods
solution to a differential equation
4.1: Basics of Differential Equations
Squeeze Theorem for Sequences
5.1: Sequences
standard form
7.5: Conic Sections
step size
4.2: Direction Fields and Numerical Methods
surface area
2.4: Arc Length of a Curve and Surface Area
Surface area generated by a parametric curve
7.2: Calculus of Parametric Curves
surface of revolution
2.4: Arc Length of a Curve and Surface Area

T
Taylor polynomials
6.3: Taylor and Maclaurin Series
Taylor series
6.3: Taylor and Maclaurin Series
Taylor's theorem with remainder
6.3: Taylor and Maclaurin Series
telescoping series
5.2: Infinite Series
term
5.1: Sequences
The Logistic Equation
4.4: The Logistic Equation
Theorem of Pappus
2.6: Moments and Centers of Mass
threshold population
4.4: The Logistic Equation
trigonometric integral
3.2: Trigonometric Integrals
trigonometric substitution
3.3: Trigonometric Substitution

V
vertex
7.5: Conic Sections
Volume by Shells
2.3: Volumes of Revolution - Cylindrical Shells

W
washer method
2.2: Determining Volumes by Slicing
Witch of Agnesi
7.1: Parametric Equations
work
2.5: Physical Applications of Integration

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: MAT 2420 Calculus II

Webpages: 63

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 84.1% (53 pages)
- Undeclared: 15.9% (10 pages)

By Page

- MAT 2420 Calculus II - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 2: Applications of Integration - *CC BY-NC-SA 4.0*
    - 2.E: Applications of Integration (Exercises) - *CC BY-NC-SA 4.0*
    - 2.0: Prelude to Applications of Integration - *CC BY-NC-SA 4.0*
    - 2.1: Areas between Curves - *CC BY-NC-SA 4.0*
    - 2.2: Determining Volumes by Slicing - *CC BY-NC-SA 4.0*
    - 2.3: Volumes of Revolution - Cylindrical Shells - *CC BY-NC-SA 4.0*
    - 2.4: Arc Length of a Curve and Surface Area - *CC BY-NC-SA 4.0*
    - 2.5: Physical Applications of Integration - *CC BY-NC-SA 4.0*
    - 2.6: Moments and Centers of Mass - *CC BY-NC-SA 4.0*
    - 2.7: Integrals, Exponential Functions, and Logarithms - *CC BY-NC-SA 4.0*
    - 2.8: Exponential Growth and Decay - *CC BY-NC-SA 4.0*
    - 2.9: Calculus of the Hyperbolic Functions - *CC BY-NC-SA 4.0*
  - 3: Techniques of Integration - *CC BY-NC-SA 4.0*
    - 3.E: Techniques of Integration (Exercises) - *CC BY-NC-SA 4.0*
    - 3.0: Prelude to Techniques of Integration - *CC BY-NC-SA 4.0*
    - 3.1: Integration by Parts - *CC BY-NC-SA 4.0*
    - 3.2: Trigonometric Integrals - *CC BY-NC-SA 4.0*
    - 3.3: Trigonometric Substitution - *CC BY-NC-SA 4.0*
    - 3.4: Partial Fractions - *CC BY-NC-SA 4.0*
    - 3.5: Other Strategies for Integration - *CC BY-NC-SA 4.0*
    - 3.6: Numerical Integration - *CC BY-NC-SA 4.0*
    - 3.7: Improper Integrals - *CC BY-NC-SA 4.0*
  - 4: Introduction to Differential Equations - *CC BY-NC-SA 4.0*
    - 4.E: Introduction to Differential Equations (Exercises) - *CC BY-NC-SA 4.0*
    - 4.0: Prelude to Differential Equations - *CC BY-NC-SA 4.0*
    - 4.1: Basics of Differential Equations - *CC BY-NC-SA 4.0*
    - 4.2: Direction Fields and Numerical Methods - *CC BY-NC-SA 4.0*
    - 4.3: Separable Equations - *CC BY-NC-SA 4.0*
    - 4.4: The Logistic Equation - *CC BY-NC-SA 4.0*
  - 5: Sequences and Series - *CC BY-NC-SA 4.0*
    - 5.E: Sequences and Series (Exercises) - *CC BY-NC-SA 4.0*
    - 5.0: Prelude to Sequence and Series - *CC BY-NC-SA 4.0*
    - 5.1: Sequences - *CC BY-NC-SA 4.0*
    - 5.2: Infinite Series - *CC BY-NC-SA 4.0*
    - 5.3: The Divergence and Integral Tests - *CC BY-NC-SA 4.0*
    - 5.4: Comparison Tests - *CC BY-NC-SA 4.0*
    - 5.5: Alternating Series - *CC BY-NC-SA 4.0*
    - 5.6: Ratio and Root Tests - *CC BY-NC-SA 4.0*
  - 6: Power Series - *CC BY-NC-SA 4.0*
    - 6.E: Power Series (Exercises) - *CC BY-NC-SA 4.0*
    - 6.0: Prelude to Power Series - *CC BY-NC-SA 4.0*
    - 6.1: Power Series and Functions - *CC BY-NC-SA 4.0*
    - 6.2: Properties of Power Series - *CC BY-NC-SA 4.0*
    - 6.3: Taylor and Maclaurin Series - *CC BY-NC-SA 4.0*
    - 6.4: Working with Taylor Series - *CC BY-NC-SA 4.0*
  - 7: Parametric Equations and Polar Coordinates - *CC BY-NC-SA 4.0*

- 7.E: Parametric Equations and Polar Coordinates (Exercises) - *CC BY-NC-SA 4.0*
- 7.0: Prelude to Parametric Equations and Polar Coordinates - *CC BY-NC-SA 4.0*
- 7.1: Parametric Equations - *CC BY-NC-SA 4.0*
- 7.2: Calculus of Parametric Curves - *CC BY-NC-SA 4.0*
- 7.3: Polar Coordinates - *CC BY-NC-SA 4.0*
- 7.4: Area and Arc Length in Polar Coordinates - *CC BY-NC-SA 4.0*
- 7.5: Conic Sections - *CC BY-NC-SA 4.0*

- Back Matter - *Undeclared*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*